

**SEARCH REQUEST FORM**

Scientific and Technical Information Center

Requester's Full Name: \_\_\_\_\_ Examiner #: \_\_\_\_\_ Date: \_\_\_\_\_  
 Art Unit: \_\_\_\_\_ Phone Number 30 \_\_\_\_\_ Serial Number: \_\_\_\_\_  
 Mail Box and Bldg/Room Location: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

**If more than one search is submitted, please prioritize searches in order of need.**

\*\*\*\*\*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

*\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

\*\*\*\*\*

**STAFF USE ONLY****Type of Search****Vendors and cost where applicable**

Searcher: _____	NA Sequence (#) _____	STN _____
Searcher Phone #: _____	AA Sequence (#) _____	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: _____	Bibliographic _____	Dr.Link _____
Date Completed: _____	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems _____
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: _____	Other _____	Other (specify) _____

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2001, 10:16:55 : Search time 41.8 Seconds  
(without alignments)  
62.964 Million cell updates/sec

Title: US-09-362-731-3

Perfect score: 733

Sequence: 1 DQYKANSKFIGITELGGQY.....FGGCHGSEPCIIHRGKPFGR 137

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PCUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	12.6	61	2	US-08-482-142-162
2	100	12.6	61	2	US-08-478-572-162
3	100	12.6	84	2	US-08-482-142-161
4	100	12.6	84	2	US-08-478-572-161
5	100	12.6	92	2	US-08-482-142-160
6	100	12.6	92	2	US-08-478-572-160
7	100	12.6	129	2	US-08-482-142-157
8	100	12.6	129	2	US-08-482-142-159
9	100	12.6	129	2	US-08-478-572-157
10	100	12.6	129	2	US-08-478-572-159
11	100	12.6	130	2	US-08-482-142-158
12	100	12.6	130	2	US-08-478-572-158
13	99	12.5	25	2	US-08-482-142-42
14	99	12.5	25	2	US-08-478-572-42
15	99	12.5	25	5	PCT-US95-04481-19
16	99	12.5	35	3	US-08-460-040-7
17	99	12.5	129	1	US-07-945-288-12
18	99	12.5	129	1	US-08-462-831-12
19	99	12.5	129	1	US-08-461-809-12
20	99	12.5	129	1	US-08-461-441-12
21	99	12.5	129	5	PCT-US93-08518-12
22	99	12.5	145	3	US-08-460-040-6
23	99	12.5	146	1	US-07-945-288-4
24	99	12.5	146	1	US-08-462-831-4
25	99	12.5	146	1	US-08-461-809-4
26	99	12.5	146	1	US-08-461-441-4
27	99	12.5	146	2	US-08-482-142-4

28	99	12.5	146	2	US-08-478-572-4	Sequence 4, Appli
29	99	12.5	146	5	PCT-US93-08518-4	Sequence 4, Appli
30	97	12.2	129	1	US-07-945-288-8	Sequence 8, Appli
31	97	12.2	129	1	US-07-945-288-13	Sequence 13, Appl
32	97	12.2	129	1	US-08-462-831-8	Sequence 8, Appli
33	97	12.2	129	1	US-08-462-831-13	Sequence 13, Appl
34	97	12.2	129	1	US-08-461-809-8	Sequence 8, Appli
35	97	12.2	129	1	US-08-461-809-13	Sequence 13, Appl
36	97	12.2	129	1	US-08-461-441-8	Sequence 8, Appli
37	97	12.2	129	1	US-08-461-441-13	Sequence 13, Appl
38	97	12.2	129	2	US-08-482-142-8	Sequence 8, Appli
39	97	12.2	129	2	US-08-478-572-8	Sequence 8, Appli
40	97	12.2	129	4	US-08-930-264-2	Sequence 2, Appli
41	97	12.2	129	4	US-08-930-264-4	Sequence 4, Appli
42	97	12.2	129	4	US-08-930-264-6	Sequence 6, Appli
43	97	12.2	129	4	US-08-930-264-8	Sequence 8, Appli
44	97	12.2	129	4	US-08-930-264-10	Sequence 10, Appl
45	97	12.2	129	5	PCT-US93-08518-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1  
US-08-482-142-162  
: Sequence 162, Application US/08482142  
: Patent No. 5820862  
: GENERAL INFORMATION:  
: APPLICANT: Garman, Richard  
: APPLICANT: Greenstein, Julia  
: APPLICANT: Kuo, Mel-chang  
: APPLICANT: Rogers, Bruce  
: APPLICANT: Franzen, Henry  
: APPLICANT: Chen, Xian  
: APPLICANT: Evans, Sean  
: APPLICANT: Shaked, Ze'ev  
: TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
: NUMBER OF SEQUENCES: 207  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
: STREET: 610 LINCOLN STREET  
: CITY: WALTHAM  
: STATE: MA  
: COUNTRY: USA  
: ZIP: 02154  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC Compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: ASCII TEXT  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/482.142  
: FILING DATE: 07-JUN-1995  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US/08/445.307  
: FILING DATE: 07 June 1995  
: ATTORNEY/AGENT INFORMATION:  
: NAME: CRAIG, ANNE I.  
: REGISTRATION NUMBER: 32,976  
: REFERENCE/DOCKET NUMBER: 017.6US  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (617) 466-6000  
: TELEFAX: (617) 466-6040  
: INFORMATION FOR SEQ ID NO: 162:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 61 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear  
: MOLECULE TYPE: peptide  
: FRAGMENT TYPE: internal  
: US-08-482-142-162

Query Match 12.6%; Score 100; DB 2; Length 61;  
 Best Local Similarity 94.1%; Pred. No. 8.4e-05;  
 Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 52 GCHGSEPCIIHRGKPF 68  
 Db 20 GCHGSEPCIIHRGKPF 36

## RESULT 2

US-08-478-572-162  
 ; Sequence 162, Application US/08478572  
 ; Patent No. 5968526  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Garman, Richard  
 ; APPLICANT: Greenstein, Julia  
 ; APPLICANT: Kuo, Mei-chang  
 ; APPLICANT: Rogers, Bruce  
 ; APPLICANT: Franzen, Henry  
 ; APPLICANT: Chen, Xian  
 ; APPLICANT: Evans, Sean  
 ; APPLICANT: Shaked, Ze'ev  
 ; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
 ; NUMBER OF SEQUENCES: 207  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
 ; STREET: 610 LINCOLN STREET  
 ; CITY: WALTHAM  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02154  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: ASCII TEXT  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/478,572  
 ; FILING DATE: 07-June-1995  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/445,307  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: CRAIG, ANNE I.  
 ; REGISTRATION NUMBER: 32,976  
 ; REFERENCE/DOCKET NUMBER: 017.605  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617) 466-6000  
 ; TELEFAX: (617) 466-6040  
 ; INFORMATION FOR SEQ ID NO: 162:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 61 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; FRAGMENT TYPE: internal  
 ; US-08-478-572-162

Query Match 12.6%; Score 100; DB 2; Length 61;  
 Best Local Similarity 94.1%; Pred. No. 8.4e-05;  
 Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 52 GCHGSEPCIIHRGKPF 68  
 Db 20 GCHGSEPCIIHRGKPF 36

## RESULT 3

US-08-482-142-161  
 ; Sequence 161, Application US/08482142  
 ; Patent No. 5820862  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Garman, Richard  
 ; APPLICANT: Greenstein, Julia  
 ; APPLICANT: Kuo, Mei-chang  
 ; APPLICANT: Rogers, Bruce  
 ; APPLICANT: Franzen, Henry  
 ; APPLICANT: Chen, Xian  
 ; APPLICANT: Evans, Sean  
 ; APPLICANT: Shaked, Ze'ev  
 ; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
 ; NUMBER OF SEQUENCES: 207  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
 ; STREET: 610 LINCOLN STREET  
 ; CITY: WALTHAM  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02154  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: ASCII TEXT  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/482,142  
 ; FILING DATE: 07-JUN-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/445,307  
 ; FILING DATE: 07 June 1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: CRAIG, ANNE I.  
 ; REGISTRATION NUMBER: 32,976  
 ; REFERENCE/DOCKET NUMBER: 017.605  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617) 466-6000  
 ; TELEFAX: (617) 466-6040  
 ; INFORMATION FOR SEQ ID NO: 161:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 84 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; FRAGMENT TYPE: internal  
 ; US-08-482-142-161

Query Match 12.6%; Score 100; DB 2; Length 84;  
 Best Local Similarity 94.1%; Pred. No. 0.00012;  
 Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 52 GCHGSEPCIIHRGKPF 68  
 Db 20 GCHGSEPCIIHRGKPF 36

## RESULT 4

US-08-478-572-161  
 ; Sequence 161, Application US/08478572  
 ; Patent No. 5968526  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Garman, Richard  
 ; APPLICANT: Greenstein, Julia  
 ; APPLICANT: Kuo, Mei-chang  
 ; APPLICANT: Rogers, Bruce  
 ; APPLICANT: Franzen, Henry  
 ; APPLICANT: Chen, Xian  
 ; APPLICANT: Evans, Sean  
 ; APPLICANT: Shaked, Ze'ev

;; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
;; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
;; NUMBER OF SEQUENCES: 207  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
;; STREET: 610 LINCOLN STREET  
;; CITY: WALTHAM  
;; STATE: MA  
;; COUNTRY: USA  
;; ZIP: 02154  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: ASCII TEXT  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/478,572  
;; FILING DATE: 07-June-1995  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/445,307  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: CRAIG, ANNE I.  
;; REGISTRATION NUMBER: 32,976  
;; REFERENCE/DOCKET NUMBER: 017.60S  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 466-6000  
;; TELEFAX: (617) 466-6040  
;; INFORMATION FOR SEQ ID NO: 161:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 84 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FRAGMENT TYPE: internal  
;; US-08-478-572-161

Query Match 12.6%; Score 100; DB 2; Length 84;  
Best Local Similarity 94.1%; Pred. No. 0.00012;  
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 52 GCHGSEPCIHRGKPKFS 68  
Db 20 GCHGSEPCIHRGKPKFT 36

## RESULT 5

US-08-482-142-160  
; Sequence 160, Application US/08482142  
; Patent No. 5820862  
; GENERAL INFORMATION:  
; APPLICANT: Garman, Richard  
; APPLICANT: Greenstein, Julia  
; APPLICANT: Kuo, Mei-chang  
; APPLICANT: Rogers, Bruce  
; APPLICANT: Franzen, Henry  
; APPLICANT: Chen, Xian  
; APPLICANT: Evans, Sean  
; APPLICANT: Shaked, Ze'ev  
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
; STREET: 610 LINCOLN STREET  
; CITY: WALTHAM  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154  
; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: ASCII TEXT  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/482,142  
;; FILING DATE: 07-JUN-1995  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/445,307  
;; FILING DATE: 07 June 1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: CRAIG, ANNE I.  
;; REGISTRATION NUMBER: 32,976  
;; REFERENCE/DOCKET NUMBER: 017.60S  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 466-6000  
;; TELEFAX: (617) 466-6040  
;; INFORMATION FOR SEQ ID NO: 160:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 92 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FRAGMENT TYPE: internal  
;; US-08-482-142-160

Query Match 12.6%; Score 100; DB 2; Length 92;  
Best Local Similarity 94.1%; Pred. No. 0.00014;  
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 52 GCHGSEPCIHRGKPKFS 68  
Db 20 GCHGSEPCIHRGKPKFT 36

## RESULT 6

US-08-478-572-160  
; Sequence 160, Application US/08478572  
; Patent No. 5968526  
; GENERAL INFORMATION:  
; APPLICANT: Garman, Richard  
; APPLICANT: Greenstein, Julia  
; APPLICANT: Kuo, Mei-chang  
; APPLICANT: Rogers, Bruce  
; APPLICANT: Franzen, Henry  
; APPLICANT: Chen, Xian  
; APPLICANT: Evans, Sean  
; APPLICANT: Shaked, Ze'ev  
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
; STREET: 610 LINCOLN STREET  
; CITY: WALTHAM  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII TEXT  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/478,572  
; FILING DATE: 07-June-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/445,307  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CRAIG, ANNE I.



REGISTRATION NUMBER: 32,976  
REFERENCE/DOCKET NUMBER: 017.60S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 466-6000  
TELEFAX: (617) 466-6040  
INFORMATION FOR SEQ ID NO: 160:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 92 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-478-572-160

Query Match 12.6%; Score 100; DB 2; Length 92;  
Best Local Similarity 94.1%; Pred. No. 0.00014;  
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 52 GCHGSEPCIHRGKPF 68  
Db 20 GCHGSEPCIHRGKPF 36

## RESULT 7

US-08-482-142-157  
Sequence 157, Application US/08482142  
Patent No. 5820862  
GENERAL INFORMATION:  
APPLICANT: Garman, Richard  
APPLICANT: Greenstein, Julia  
APPLICANT: Kuo, Mei-chang  
APPLICANT: Rogers, Bruce  
APPLICANT: Franzen, Henry  
APPLICANT: Chen, Xian  
APPLICANT: Evans, Sean  
APPLICANT: Shaked, Ze'ev  
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
OPERATING SYSTEM: PC-DOS/MS-DOS  
NUMBER OF SEQUENCES: 207  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
STREET: 610 LINCOLN STREET  
CITY: WALTHAM  
STATE: MA  
COUNTRY: USA  
ZIP: 02154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,142  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/445,307  
FILING DATE: 07 June 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: CRAIG, ANNE I.  
REGISTRATION NUMBER: 32,976  
REFERENCE/DOCKET NUMBER: 017.60S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 466-6000  
TELEFAX: (617) 466-6040  
INFORMATION FOR SEQ ID NO: 157:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 129 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal

US-08-482-142-157

Query Match 12.6%; Score 100; DB 2; Length 129;  
Best Local Similarity 94.1%; Pred. No. 0.0002;  
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 52 GCHGSEPCIHRGKPF 68  
Db 20 GCHGSEPCIHRGKPF 36

## RESULT 8

US-08-482-142-159  
Sequence 159, Application US/08482142  
Patent No. 5820862  
GENERAL INFORMATION:  
APPLICANT: Garman, Richard  
APPLICANT: Greenstein, Julia  
APPLICANT: Kuo, Mei-chang  
APPLICANT: Rogers, Bruce  
APPLICANT: Franzen, Henry  
APPLICANT: Chen, Xian  
APPLICANT: Evans, Sean  
APPLICANT: Shaked, Ze'ev  
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
OPERATING SYSTEM: PC-DOS/MS-DOS  
NUMBER OF SEQUENCES: 207  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
STREET: 610 LINCOLN STREET  
CITY: WALTHAM  
STATE: MA  
COUNTRY: USA  
ZIP: 02154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,142  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/445,307  
FILING DATE: 07 June 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: CRAIG, ANNE I.  
REGISTRATION NUMBER: 32,976  
REFERENCE/DOCKET NUMBER: 017.60S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 466-6000  
TELEFAX: (617) 466-6040  
INFORMATION FOR SEQ ID NO: 159:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 129 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-482-142-159

Query Match 12.6%; Score 100; DB 2; Length 129;  
Best Local Similarity 94.1%; Pred. No. 0.0002;  
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 52 GCHGSEPCIHRGKPF 68  
Db 20 GCHGSEPCIHRGKPF 36

## RESULT 9

US-08-478-572-157  
; Sequence 157, Application US/08478572  
; Patent No. 5968526  
; GENERAL INFORMATION:  
; APPLICANT: Garman, Richard  
; APPLICANT: Greenstein, Julia  
; APPLICANT: Kuo, Mei-chang  
; APPLICANT: Rogers, Bruce  
; APPLICANT: Franzen, Henry  
; APPLICANT: Chen, Xian  
; APPLICANT: Evans, Sean  
; APPLICANT: Shaked, Ze'ev  
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
; STREET: 610 LINCOLN STREET  
; CITY: WALTHAM  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII TEXT  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/478, 572  
; FILING DATE: 07-June-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/445, 307  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CRAIG, ANNE I.  
; REGISTRATION NUMBER: 32,976  
; REFERENCE/DOCKET NUMBER: 017.6US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 466-6000  
; TELEFAX: (617) 466-6040  
; INFORMATION FOR SEQ ID NO: 157:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 129 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
US-08-478-572-157

Query Match 12.6%; Score 100; DB 2; Length 129;  
Best Local Similarity 94.1%; Pred. No. 0.0002;  
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 52 GCHGSEPCIHRGKPPS 68  
DB 20 GCHGSEPCIHRGKPPFT 36  
US-08-478-572-157

## RESULT 10

US-08-478-572-159  
; Sequence 159, Application US/08478572  
; Patent No. 5968526  
; GENERAL INFORMATION:  
; APPLICANT: Garman, Richard  
; APPLICANT: Greenstein, Julia  
; APPLICANT: Kuo, Mei-chang  
; APPLICANT: Rogers, Bruce  
; APPLICANT: Franzen, Henry  
; APPLICANT: Chen, Xian  
; APPLICANT: Evans, Sean

; APPLICANT: Shaked, Ze'ev  
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
; STREET: 610 LINCOLN STREET  
; CITY: WALTHAM  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII TEXT  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/478, 572  
; FILING DATE: 07-June-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/445, 307  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CRAIG, ANNE I.  
; REGISTRATION NUMBER: 32,976  
; REFERENCE/DOCKET NUMBER: 017.6US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 466-6000  
; TELEFAX: (617) 466-6040  
; INFORMATION FOR SEQ ID NO: 159:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 129 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
US-08-478-572-159

Query Match 12.6%; Score 100; DB 2; Length 129;  
Best Local Similarity 94.1%; Pred. No. 0.0002;  
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 52 GCHGSEPCIHRGKPPS 68  
DB 20 GCHGSEPCIHRGKPPFT 36  
US-08-478-572-159

## RESULT 11

US-08-482-142-158  
; Sequence 158, Application US/08482142  
; Patent No. 5820862  
; GENERAL INFORMATION:  
; APPLICANT: Garman, Richard  
; APPLICANT: Greenstein, Julia  
; APPLICANT: Kuo, Mei-chang  
; APPLICANT: Rogers, Bruce  
; APPLICANT: Franzen, Henry  
; APPLICANT: Chen, Xian  
; APPLICANT: Evans, Sean  
; APPLICANT: Shaked, Ze'ev  
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
; STREET: 610 LINCOLN STREET  
; CITY: WALTHAM  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154  
; COMPUTER READABLE FORM:

```

: NAME: CRAIG, ANNE I.
: REGISTRATION NUMBER: 32,976
: REFERENCE/DOCKET NUMBER: 017.6US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 466-6000
: TELEFAX: (617) 466-6040
: INFORMATION FOR SEQ ID NO: 158:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 130 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: FRAGMENT TYPE: internal
: US-08-478-572-158

Query Match 12.68; Score 100; DB 2: Length 130;
Best Local Similarity 94.1%; Pred. No. 0.0002;
Matches 16; Conservative 1; Mismatches 0; Indels 0;

QY 52 GCGSEPCIIHRGKPF 68
      |||||
Db 20 GCGSEPCIIHRGPFT 36

RESULT 13
US-08-482-142-42
: Sequence 42, Application US/08482142
: Patent No. 5820862
: GENERAL INFORMATION:
: APPLICANT: Garman, Richard
: APPLICANT: Greenstein, Julia
: APPLICANT: Kuo, Mei-chang
: APPLICANT: Rogers, Bruce
: APPLICANT: Franzen, Henry
: APPLICANT: Chen, Xian
: APPLICANT: Evans, Sean
: APPLICANT: Shaked, Ze'ev
: TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
: TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
: NUMBER OF SEQUENCES: 207
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
: STREET: 610 LINCOLN STREET
: CITY: WALTHAM
: STATE: MA
: COUNTRY: USA
: ZIP: 02154
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII TEXT
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/482,142
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/445,307
: FILING DATE: 07 June 1995
: ATTORNEY/AGENT INFORMATION:
: NAME: CRAIG, ANNE I.
: REGISTRATION NUMBER: 32,976
: REFERENCE/DOCKET NUMBER: 017.6US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 466-6000
: TELEFAX: (617) 466-6040
: INFORMATION FOR SEQ ID NO: 42:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 25 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide

```

Db 10 GCHGSEPCIIHRGKPF 25

Db 10 GCHGSEPCIIHRGKPF 25

Search completed: April 14, 2001, 10:16:55  
Job time: 440 sec

Qy 52 GCHGSEPCIHRGKPF 67  
 |||||  
 Db 10 GCHGSEPCIHRGKPF 25

Qy 52 GCHGSEPCIHRGKPF 67  
 |||||  
 Db 10 GCHGSEPCIHRGKPF 25

QY	52	GCHGSEPCIIHRGKPF	67
Db	10	GCHGSEPCIIHRGKPF	25



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2001, 10:17:44 ; Search time 27.31 Seconds  
(without alignments)  
40.138 Million cell updates/sec

Title: US-09-362-731-5

Perfect score: 181

Sequence: 1 QYIRANSKFIGITELGGCHGSEPCNIHRKPF 32

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	50.8	146	1	DER2_DERPT
2	89	49.2	146	1	DEF2_DERFA
3	74	40.9	1314	1	TETX_CLOTE
4	51	28.2	307	1	YD94_HELPJ
5	49	27.1	238	1	IBP6_MOUSE
6	49	27.1	259	1	LPXA_CHRVI
7	48.5	26.8	378	1	URH1_YEAST
8	48	26.5	612	1	GIDA_MYCGE
9	47	26.0	398	1	RFE_MYCLE
10	47	26.0	404	1	RFE_MYCTU
11	47	26.0	406	1	BHMT_HUMAN
12	47	26.0	407	1	BHMT_MOUSE
13	47	26.0	407	1	BHMT_RAT
14	46.5	25.7	141	1	ALG2_TYRPU
15	46.5	25.7	357	1	METX_HAEIN
16	46.5	25.7	379	1	METX_NEIMA
17	46.5	25.7	379	1	METX_NEIMB
18	46	25.4	339	1	TCMO_STRGA
19	46	25.4	379	1	METX_LEPME
20	45.5	25.1	141	1	LEPI_LEPDS
21	45.5	25.1	470	1	NRAM_IARUE
22	45.5	25.1	470	1	NRAM_IATRA
23	45.5	25.1	470	1	NRAM_IAWHM
24	45.5	25.1	4393	1	PCBM_HUMAN
25	45	24.9	301	1	SPRC_RAT
26	45	24.9	302	1	SPRC_MOUSE
27	45	24.9	382	1	METX_MYCLE
28	45	24.9	521	1	COAT_BPT4
29	45	24.9	773	1	DPOL_THEGO
30	45	24.9	780	1	ZOR4_XENLA
31	44.5	24.6	125	1	LEFB_NPVOP
32	44.5	24.6	380	1	METX_THETH
33	44.5	24.6	397	1	TDG_MOUSE

34	44.5	24.6	2156	1	ORP1_HUMAN
35	44.5	24.6	2214	1	POLG_CXA24
36	44.5	24.6	3707	1	PGEM_MOUSE
37	44	24.3	66	1	VG84_BPMLS
38	44	24.3	195	1	GYRA_FIBSU
39	44	24.3	256	1	YD83_METJA
40	44	24.3	321	1	DHOA_NEUCR
41	44	24.3	329	1	DHOA_EMENI
42	44	24.3	379	1	METX_MYCTU
43	44	24.3	532	1	SREP_PENCH
44	44	24.3	535	1	2257_HUMAN
45	44	24.3	619	1	YOL8_CAEEL

## ALIGNMENTS

### RESULT 1

DER2\_DERPT  
ID DER2\_DERPT STANDARD; PRT; 146 AA.  
AC P49278;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE MITE ALLERGEN DER P 2 PRECURSOR (DER P II) (DPX).  
GN DERP2.  
OS Dermatophagoides pteronyssinus (House-dust mite).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
OC Acariformes; Sarcoptiformes; Astigmata; Analgoidea; Pyroglyphidae;  
OC Dermatophagoides.  
OX NCBI\_TaxID=6956;  
RX MEDLINE=90256301; PubMed=2341191;  
RA Chua K.Y., Doyle C.R., Simpson R.J., Turner K.J., Stewart G.A.,  
RA Thomas W.R.;  
RT "Isolation of cDNA coding for the major mite allergen Der p II by IgE  
plaque immunoassay."  
RL Int. Arch. Allergy Appl. Immunol. 91:118-123(1990).  
RN [2]  
RP PARTIAL SEQUENCE OF 18-57.  
RX MEDLINE=89278484; PubMed=2732406;  
RA Heymann P.W., Chapman M.D., Aalberse R.C., Fox J.W.,  
RA Platts-Mills T.A.;  
RT "Antigenic and structural analysis of group II allergens (Der f II  
and Der p II) from house dust mites (Dermatophagoides spp).";  
RL J. Allergy Clin. Immunol. 83:1055-1067(1989).  
RN [3]  
RP VARIANTS.  
RA Smith W., Hales B.J., Thomas W.R.;  
RT "Allergens of wild house dust mites: environmental Der p 1 and Der p 2  
sequence polymorphisms";  
RL Submitted (JUN-2000) to the SWISS-PROT data bank.  
RN [4]  
RP STRUCTURE BY NMR.  
RX MEDLINE=98409423; PubMed=9737847;  
RA Mueller G.A., Benjamin D.C., Rule G.S.;  
RT "tertiary structure of the major house dust mite allergen Der p 2:  
sequential and structural homologies";  
RL Biochemistry 37:12707-12714(1998).  
CC -!- SIMILARITY: TO MITE ALLERGEN LEP D I.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; AF276239; AAF86462.1; -  
DR PDB; 1A9V; 14-OCT-98.  
KW Allergen; Signal; 3D-structure; Polymorphism.

```
FT SIGNAL 1 17
FT CHAIN 18 146
FT DISULFID 25 136
FT DISULFID 38 44
FT DISULFID 90 95
FT VARIANT 39 39
FT VARIANT 40 40
FT VARIANT 44 44
FT VARIANT 47 47
FT VARIANT 49 49
FT VARIANT 56 56
FT VARIANT 61 61
FT VARIANT 75 75
FT VARIANT 78 78
FT VARIANT 81 81
FT VARIANT 95 95
FT VARIANT 98 98
FT VARIANT 108 108
FT VARIANT 111 111
FT VARIANT 114 114
FT VARIANT 116 116
FT VARIANT 118 118
FT VARIANT 127 127
SQ SEQUENCE 146 AA; 15999 MW; 591B2FA7FD26D3AF CRC64;

Query Match 50.8%; Score 92; DB 1; Length 146;
Best Local Similarity 93.8%; Pred. No. 1.9e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 17 GCHGSEPCNIHRGKPF 32
Db 37 GCHGSEPCNIHRGKPF 52

RESULT 2
DEF2_DERFA STANDARD; PRT; 146 AA.
AC Q00855; P39672; Q26359;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MITE ALLERGEN DER F 2 PRECURSOR (DER F II).
GN DERF2.
OS Dermatophagoides farinae (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcopitiformes; Astigmata; Analgoidea; Pyroglyphidae;
OC Dermatophagoides.
OX NCBI_TaxID=6954;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=91291341; PubMed=1368682;
RA Yuuki T., Okumura Y., Ando T., Yamakawa H., Suko M., Haida M.,
RA Okudaira H.;
RT "Cloning and expression of cDNA coding for the major house dust mite
RT allergen Der f II in Escherichia coli.";
RL Agric. Biol. Chem. 55:1233-1238(1991).
RN [2]
RP SEQUENCE OF 4-146 FROM N.A.
RX MEDLINE=94256850; PubMed=8198452;
RA Okuhira H.;
RT "Molecular biology of mite antigens.";
RL Arerugi 43:435-440(1994).
RN [3]
RP DISULFIDE BONDS, AND PARTIAL SEQUENCE.
RX MEDLINE=93283958; PubMed=8508052;
RA Nishiyama C., Yuuki T., Takai T., Okumura Y., Okudaira H.;
RT "Determination of three disulfide bonds in a major house dust mite
RT allergen, Der f II.";
RL Int. Arch. Allergy Immunol. 101:159-166(1993).
RN [4]
RP PARTIAL SEQUENCE OF 18-52.
RX MEDLINE=89278484; PubMed=2732406;
```

```
RA Heymann P.W., Chapman M.D., Aalberse R.C., Fox J.W.,
RA Platts-Mills T.A.;
RT "Antigenic and structural analysis of group II allergens (Der f II
RT and Der p II) from house dust mites (Dermatophagoides spp).";
RL J. Allergy Clin. Immunol. 83:1055-1067(1989).
RN [5]
RP STRUCTURE BY NMR.
RX MEDLINE=98079088; PubMed=9417088;
RA Ichikawa S., Hatanaka H., Yuuki T., Iwamoto N., Kojima S.,
RA Nishiyama C., Ogura K., Okumura Y., Inagaki F.;
RT "Solution structure of Der f 2, the major mite allergen for atopic
RT diseases.";
RL J. Biol. Chem. 273:356-360(1998).
CC -!- MISCELLANEOUS: THE SEQUENCE SHOWN HERE IS FROM CLONE 2. THE
CC N-TERMINAL SEQUENCE (AA 1-8) FROM CLONE 1 AND 11 ARE NOT YET
CC KNOWN.
CC -!- SIMILARITY: TO MITE ALLERGEN LEP D I.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D10447; BAA01239.1; -
CC EMBL; D10448; BAA01240.1; -
CC EMBL; D10449; BAA01241.1; -
CC EMBL; S70378; AAB30829.1; -
CC PIR; PS0417; PS0417.
CC PDB; 1AHK; 08-APR-98.
CC PDB; 1AHM; 08-APR-98.
KW Allergen; Signal; Polymorphism; 3D-structure.
FT SIGNAL 1 17
FT CHAIN 18 146 MITE ALLERGEN DER F 2.
FT DISULFID 25 136
FT DISULFID 38 44
FT DISULFID 90 95
FT VARIANT 93 93 M -> V (IN CLONE 1).
FT VARIANT 105 105 I -> A (IN CLONE 11).
FT VARIANT 128 128 I -> V (IN CLONE 11).
FT VARIANT 142 142 G -> A (IN CLONE 11).
FT CONFLICT 5 8 ILCL -> GTMV (IN REF. 2).
SQ SEQUENCE 146 AA; 15802 MW; FAL18206CD88534A CRC64;

Query Match 49.2%; Score 89; DB 1; Length 146;
Best Local Similarity 87.5%; Pred. No. 5.2e-06;
Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 17 GCHGSEPCNIHRGKPF 32
Db 37 GCHGSEPCNIHRGKPF 52

RESULT 3
TETX_CLOTE STANDARD; PRT; 1314 AA.
AC P04958;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TETANUS TOXIN PRECURSOR (EC 3.4.24.68) (TENTOXLYSIN).
OS Clostridium tetani.
OG Plasmid.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87053814; PubMed=3536478;
RA Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,
```

RA Weller U., Hudel M., Habermann E., Niemann H.:  
 RT "Tetanus toxin: primary structure, expression in *E. coli*, and  
 RL homology with botulinum toxins.";  
 RN EMBO J. 5:2495-2502(1986).  
 RP [2]  
 RC SEQUENCE FROM N.A.  
 RX STRAIN-CN3911;  
 RA MEDLINE=87040747; PubMed=3774547;  
 RT Fairweather N.F., Lyness V.A.;  
 RL "The complete nucleotide sequence of tetanus toxin.";  
 RN Nucleic Acids Res. 14:7809-7812(1986).  
 RP [3]  
 RX SEQUENCE OF 742-1314 FROM N.A.  
 RA MEDLINE=86085672; PubMed=3510187;  
 RT Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;  
 RL "Cloning, nucleotide sequencing, and expression of tetanus toxin  
 fragment C in *Escherichia coli*.";  
 RN J. Bacteriol. 165:21-27(1986).  
 RP [4]  
 RX PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RA MEDLINE=90201034; PubMed=2108021;  
 RT Krieglstein K., Henschen A., Weller U., Habermann E.;  
 RL "Arrangement of disulfide bridges and positions of sulphydryl groups  
 in tetanus toxin.";  
 RN Eur. J. Biochem. 188:39-45(1990).  
 RP [5]  
 RX PARTIAL SEQUENCE.  
 RA MEDLINE=92037649; PubMed=1935979;  
 RT Krieglstein K.G., Henschen A.H., Weller U., Habermann E.;  
 RL "Limited proteolysis of tetanus toxin. Relation to activity and  
 identification of cleavage sites.";  
 RN Eur. J. Biochem. 202:41-51(1991).  
 RP [6]  
 RX IDENTIFICATION AS ZINC-PROTEASE.  
 RA MEDLINE=93010948; PubMed=1336558;  
 RT Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,  
 RL Montecucco C.;  
 RP "Tetanus toxin is a zinc protein and its inhibition of  
 neurotransmitter release and protease activity depend on zinc.";  
 RN EMBO J. 11:3577-3583(1992).  
 RP [7]  
 RX IDENTIFICATION OF SUBSTRATE.  
 RA MEDLINE=93063293; PubMed=1331807;  
 RT Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,  
 RL Dasgupta B.R., Montecucco C.;  
 RP "Tetanus and botulinum-B neurotoxins block neurotransmitter release  
 by proteolytic cleavage of synaptobrevin.";  
 RN Nature 359:832-835(1992).  
 RP [8]  
 RX X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.  
 RA MEDLINE=97475217; PubMed=9334741;  
 RT Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,  
 RL Sax M.;  
 RP "Structure of the receptor binding fragment HC of tetanus  
 neurotoxin.";  
 RL Nat. Struct. Biol. 4:788-792(1997).  
 CC -1- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77  
 BOND OF SYNAPTOSOMALIN-2.  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-|-PHE-77 BOND IN  
 SYNAPTOSOMALIN.  
 CC -1- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO  
 YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE  
 AND ARE NON-TOXIC AFTER SEPARATION.  
 CC -1- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO  
 GLYCOSIDE RECEPTORS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC  
 METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN  
 SUBFAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X04436; CAA28033.1; -;  
 DR EMBL; M12739; AAA23282.1; -;  
 DR EMBL; X06214; CAA29564.1; -;  
 DR PIR; A25689; BTCLTN.  
 DR PDB; 1AF9; 29-APR-98.  
 DR PDB; 1A8D; 14-OCT-98.  
 DR MEROPS; M27.001; -;  
 DR InterPro; IPR000130; -;  
 DR InterPro; IPR000395; -;  
 DR Pfam; PF01742; Peptidase\_M27; 1.  
 DR PRINTS; PR00760; BONTOXILYSIN.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;  
 KW 3D-structure.  
 FT INIT\_MET 0  
 FT CHAIN 457 1 456  
 FT CHAIN 457 1314  
 FT METAL 232 232  
 FT ACT\_SITE 233 233  
 FT METAL 236 236  
 FT TRANSMEM 226 246  
 FT TRANSMEM 669 689  
 FT DISULFID 438 466  
 FT DISULFID 1076 1092  
 FT SEQUENCE 1314 AA; 150550 MW; 134C3657133EF81D CRC64;  
 SQ  
 Query Match 40.9%; Score 74; DB 1; Length 1314;  
 Best Local Similarity 100.0%; Pred. No. 0.0064;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 QYIKANSKFIGITEL 15  
 Db 829 QYIKANSKFIGITEL 843  
 RESULT 4  
 YD94\_HELPJ STANDARD; PRT; 307 AA.  
 AC Q9ZJ81;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE HYPOTHETICAL PROTEIN JHP1433.  
 GN JHP1433.  
 OS Helicobacter pylori J99 (Campylobacter pylori J99).  
 OC Bacteria; proteobacteria; epsilon subdivision; Helicobacter group;  
 OC Helicobacter.  
 OX NCBI\_TaxID=85963;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95120557; PubMed=9923682;  
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,  
 RA Smith D.R., Noonan B., Gullid B.C., deJonge B.L., Carmel G.,  
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,  
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,  
 RA Trust T.J.;  
 RT "Genomic sequence comparison of two unrelated isolates of the human  
 RT gastric pathogen *Helicobacter pylori*.";  
 RL Nature 397:176-180(1999).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0119 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AE001566; AAD07009.1; -  
 DR Pfam; PF01513; DUF15; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 307 AA; 34158 MW; 6DF648F20BB3D94F CRC64;

Query Match 28.2%; Score 51; DB 1; Length 307;  
 Best Local Similarity 27.6%; Pred. No. 3.6;  
 Matches 8; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITELGGCHGSEPCNIHRG 29  
 Db 163 EIVIAKKALGVLDIKACAGHTPFNTYRG 191

RESULT 5  
 IBP6\_MOUSE  
 ID IBP6\_MOUSE STANDARD; PRT; 238 AA.  
 AC P47880;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 6 PRECURSOR (IGFBP-6)  
 DE (IBP-6) (IGF-BINDING PROTEIN 6).  
 GN IGFBP6 OR IGFBP-6.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=95121750; PubMed=7529732;  
 RA Schuller A.G.P., Groffen C., van Neck J.W., Zwarthoff E.C.,  
 Drop S.L.S.;  
 RT "cDNA cloning and mRNA expression of the six mouse insulin-like  
 RT growth factor binding proteins.";  
 RL Mol. Cell. Endocrinol. 104:57-66(1994).  
 CC -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFs  
 CC AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH  
 CC PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE  
 CC INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS.  
 CC -!- SUBCELLULAR LOCATION: SECRETED.  
 CC -!- PTM: O-GLYCOSYLATED.  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
 CC -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
 CC PROTEIN FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X81584; CAA57274.1; -  
 DR HSP; P18619; IFVL.  
 DR MGD; MG1:96441; Igfbp6.  
 DR InterPro; IPR000716; -  
 DR InterPro; IPR000867; -  
 DR Pfam; PF00219; IGFBP; 1.  
 DR Pfam; PF00086; thyroglobulin\_1; 1.  
 DR PROSITE; PS00222; IGF\_BINDING; FALSE\_NEG.  
 DR PROSITE; PS00484; THYROGLOBULIN\_1; 1.  
 KW Growth factor binding; Signal; Glycoprotein.  
 FT SIGNAL 1 25 BY SIMILARITY.

FT CHAIN 26 238 INSULIN-LIKE GROWTH FACTOR BINDING  
 FT DOMAIN 184 232 PROTEIN 6.  
 FT DISULFID 30 33 THYROGLOBULIN TYPE I.  
 FT DISULFID 41 45 BY SIMILARITY.  
 FT DISULFID 58 64 BY SIMILARITY.  
 FT DISULFID 72 85 BY SIMILARITY.  
 FT DISULFID 79 105 BY SIMILARITY.  
 FT DISULFID 160 188 BY SIMILARITY.  
 FT DISULFID 199 210 BY SIMILARITY.  
 FT DISULFID 212 232 BY SIMILARITY.  
 SQ SEQUENCE 238 AA; 25402 MW; CCEDC1D6AD9D59F9 CRC64;

Query Match 27.1%; Score 49; DB 1; Length 238;  
 Best Local Similarity 39.3%; Pred. No. 5.6;  
 Matches 11; Conservative 3; Mismatches 12; Indels 2; Gaps 1;

QY 5 ANSKFIGITELGGC--HGSEPCNIHRGK 30  
 Db 51 AGSPADGCTEAGGLRRREGQPCGVSPK 78

RESULT 6  
 LPXA\_CHRVI  
 ID LPXA\_CHRVI STANDARD; PRT; 259 AA.  
 AC Q46481;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 35, Last annotation update)  
 DE ACYL-[ACYL-CARRIER-PROTEIN]-UDP-N-ACETYLGLUCOSAMINE O-ACYLTRANSFERASE  
 DE (EC 2.3.1.129) (UDP-N-ACETYLGLUCOSAMINE ACYLTRANSFERASE).  
 GN LPXA.  
 OS Chromatium vinosum.  
 CC Bacteria; Proteobacteria; gamma subdivision; Chromatiaceae;  
 CC Allochrochromatium.  
 OX NCBI\_TaxID=1049;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-D / ATCC 17899 / DSM 180;  
 RA Chen Y.L., Knaff D.B.;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF LIPID A, A  
 CC PHOSPHORYLATED GLYCOPID THAT ANCHORS THE LIPOPOLYSACCHARIDE TO  
 CC THE OUTER MEMBRANE OF THE CELL (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: (R)-3-HYDROXYTETRADECANOYL-[ACYL-CARRIER  
 CC PROTEIN] + UDP-N-ACETYLGLUCOSAMINE -> [ACYL-CARRIER PROTEIN] +  
 CC UDP-3-O-(3-HYDROXYTETRADECANOYL)-N-ACETYLGLUCOSAMINE.  
 CC -!- PATHWAY: FIRST STEP IN LIPID A BIOSYNTHESIS.  
 CC -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- SIMILARITY: BELONGS TO THE CYSE/LACA/LPXA/NOBL FAMILY OF  
 CC ACETYLTRANSFERASES. COMPOSED OF MULTIPLE REPEATS OF [LIV]-G-X(4).  
 CC -----

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; L76417; AAB02979.1; -  
 DR HSP; P10440; ILXA.  
 DR InterPro; IPR001451; -  
 DR Pfam; PF00132; hexapep; 2.  
 DR PROSITE; PS00101; HEXAPEP\_TRANSFERASE; FALSE\_NEG.  
 KW Transferase; Acyltransferase; Lipid A biosynthesis; Lipid synthesis;  
 KW Repeat.  
 SQ SEQUENCE 259 AA; 28172 MW; D7F4690066180CDA CRC64;

Query Match 27.1%; Score 49; DB 1; Length 259;

```

Best Local Similarity 37.9%; Pred. No. 6.1;
Matches 11; Conservative 5; Mismatches 5; Indels 8; Gaps 2;

QY 3 IKANSKFIGITELGG-----CH-----GSEP 23
   | : | : | : | : | : | : | : | : | : |
Db 39 IESNVRIFGVTRMGAAHNRVCHGATLGSEP 67

RESULT 7
URH1_YEAST
ID URH1_YEAST STANDARD; PRT; 378 AA.
AC Q04179;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE URIDINE NUCLEOSIDASE (EC 3.2.2.3) (URIDINE RIBOHYDROLASE).
GN URH1 OR YDRA00W OR D9509.19.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E.,
RA Berno A., Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,
RA Husek-Smith S., Hyman R., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X.,
RA Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N.,
RA Winant A., Yelton M., Bolstein D., Davis R.W.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=99431800; PubMed=10501935;
RA Kurtz J.-E., Exinger F., Erbs P., Jund R.;
RT "New insights into the pyrimidine salvage pathway of Saccharomycetes
RT cerevisiae: requirement of six genes for cytidine metabolism.";
RL Curr. Genet. 36:130-136(1999).
CC -!- FUNCTION: ALSO ACTS ON CYTIDINE.
CC -!- CATALYTIC ACTIVITY: URIDINE + H(2)O = URACIL + D-RIBOSE.
CC -!- SIMILARITY: BELONGS TO THE IUNH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32274; AAB64841.1; -
DR HSP; Q27546; IMAS.
DR SGD; S0002808; URH1.
DR InterPro; IPR001910; -
DR Pfam; PF01156; IU_nuc_hydro; 1.
DR PROSITE; PS01247; IUNH; 1.
KW Hydrolase; Glycosidase.
FT ACT_SITE 292
FT ACT_SITE 292 BY SIMILARITY.
SQ SEQUENCE 378 AA; 42317 MW; 58499E2B7480BEE5 CRC64;

Query Match 26.8%; Score 48.5; DB 1; Length 378;
Best Local Similarity 45.8%; Pred. No. 10;
Matches 11; Conservative 3; Mismatches 9; Indels 1; Gaps 1;

QY 2 YIKANSKFIGITELGGCHGSEPNCN 25
   | : | : | : | : | : | : | : | : | : |
Db 183 YLKSVKVIISIMG-GGLHGLNCN 205

RESULT 8
GIDA_MYCGE
ID GIDA_MYCGE STANDARD; PRT; 612 AA.
AC P47619;

Query Match 26.5%; Score 48; DB 1; Length 612;
Best Local Similarity 41.4%; Pred. No. 19;
Matches 12; Conservative 1; Mismatches 12; Indels 4; Gaps 1;

QY 3 IKANSKFIGITELGGCHGSEPNCNIHGRKP 31
   | : | : | : | : | : | : | : | : | : |
Db 26 IKVNLVLVDINHLSGSC----PCNPSIGGP 50

RESULT 9
RFE_MYCLE
ID RFE_MYCLE STANDARD; PRT; 398 AA.
AC P45630;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PUTATIVE UNDECAPRENYL-PHOSPHATE ALPHA-N-ACETYLGLUCOSAMINYLTRANSFERASE
DE (EC 2.4.1.-).
GN RFE.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;

```

```

DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE GLUCOSE INHIBITED DIVISION PROTEIN A.
GN GIDA OR MG379.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudak D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RL "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 301-394 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -!- FUNCTION: NOT KNOWN.
CC -!- SIMILARITY: BELONGS TO THE GIDA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39719; AAC71606.1; -
DR EMBL; U01812; AAD12347.1; -
DR TIGR; MG379; -
DR InterPro; IPR002218; -
DR Pfam; PF01134; GIDA_1; 1.
DR PROSITE; PS01280; GIDA_1; 1.
DR PROSITE; PS01281; GIDA_2; 1.
SQ SEQUENCE 612 AA; 68995 MW; 67A3F64C985B5912 CRC64;

Query Match 26.5%; Score 48; DB 1; Length 612;
Best Local Similarity 41.4%; Pred. No. 19;
Matches 12; Conservative 1; Mismatches 12; Indels 4; Gaps 1;

QY 3 IKANSKFIGITELGGCHGSEPNCNIHGRKP 31
   | : | : | : | : | : | : | : | : | : |
Db 26 IKVNLVLVDINHLSGSC----PCNPSIGGP 50

RESULT 9
RFE_MYCLE
ID RFE_MYCLE STANDARD; PRT; 398 AA.
AC P45630;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PUTATIVE UNDECAPRENYL-PHOSPHATE ALPHA-N-ACETYLGLUCOSAMINYLTRANSFERASE
DE (EC 2.4.1.-).
GN RFE.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;

```

RN SEQUENCE FROM N.A.  
 RP Smith D.R., Robinson K.;  
 RA Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 RL -!- FUNCTION: MAY BE THE TUNICAMYCIN SENSITIVE TRANSFERASE THAT  
 CC CATALYZES THE SYNTHESIS OF GLCNAC-PYROPHOSPHORYLUNDECAPRENOL  
 CC (LIPID I), THE FIRST LIPID-LINKED INTERMEDIATE INVOLVED IN BCA  
 CC SYNTHESIS (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + UNDECAPRENYL  
 CC PHOSPHATE = UMP + UNDECAPRENYL N-ACETYL-ALPHA-D-GLUCOSAMINYL  
 CC PYROPHOSPHATE.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -!- SIMILARITY: SOME, TO THE DOLICHYL-PHOSPHATE ALPHA-N-ACETYL-  
 CC GLUCOSAMINYLTRANSFERASES OF YEAST AND MAMMALS.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U15186; AAA63094.1; -.  
 CC InterPro: IPR000715; -.  
 CC Pfam: PF00953; Glycosyltransferase; Transferase;  
 KW Lipopolysaccharide biosynthesis; Glycosyltransferase; Transferase;  
 KW Transmembrane. 33 53 POTENTIAL.  
 FT TRANSMEM 79 99 POTENTIAL.  
 FT TRANSMEM 105 125 POTENTIAL.  
 FT TRANSMEM 148 168 POTENTIAL.  
 FT TRANSMEM 175 195 POTENTIAL.  
 FT TRANSMEM 198 218 POTENTIAL.  
 FT TRANSMEM 225 245 POTENTIAL.  
 FT TRANSMEM 260 280 POTENTIAL.  
 FT TRANSMEM 295 315 POTENTIAL.  
 FT TRANSMEM 347 367 POTENTIAL.  
 FT TRANSMEM 372 392 POTENTIAL.  
 SQ SEQUENCE 398 AA; 42088 MW; 37564293606CF9A6 CRC64;  
  
 Query Match 26.0%; Score 47; DB 1; Length 398;  
 Best Local Similarity 43.5%; Pred. No. 18;  
 Matches 10; Conservative 1; Mismatches 12; Indels 0; Gaps 0;  
  
 OY 10 IGITELGGCHGSEPCNIHRGKPF 32  
 | : | | | | | | | | | |  
 Db 233 ISVVLGAGCLGFLPHNFRKIP 255  
  
 RESULT 10  
 RFE\_MYCTU STANDARD; PRT; 404 AA.  
 AC Q10606;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE PUTATIVE UNDECAPRENYL-PHOSPHATE ALPHA-N-ACETYLGLUCOSAMINYLTRANSFERASE  
 DE (EC 2.4.1.-).  
 GN RFE OR RV1302 OR MTCY373.22.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 CC -!- FUNCTION: MAY BE THE TUNICAMYCIN SENSITIVE TRANSFERASE THAT  
 CC CATALYZES THE SYNTHESIS OF GLCNAC-PYROPHOSPHORYLUNDECAPRENOL  
 CC (LIPID I), THE FIRST LIPID-LINKED INTERMEDIATE INVOLVED IN BCA  
 CC SYNTHESIS (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + UNDECAPRENYL  
 CC PHOSPHATE = UMP + UNDECAPRENYL N-ACETYL-ALPHA-D-GLUCOSAMINYL  
 CC PYROPHOSPHATE.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -!- SIMILARITY: SOME, TO THE DOLICHYL-PHOSPHATE ALPHA-N-ACETYL-  
 CC GLUCOSAMINYLTRANSFERASES OF YEAST AND MAMMALS.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z734119; CA97735.1; -.  
 CC InterPro: IPR000715; -.  
 CC Pfam: PF00953; Glycosyltransferase; Glycosyltransferase; Transferase;  
 KW Lipopolysaccharide biosynthesis; Glycosyltransferase; Transferase;  
 KW Transmembrane. 3 23 POTENTIAL.  
 FT TRANSMEM 33 53 POTENTIAL.  
 FT TRANSMEM 79 99 POTENTIAL.  
 FT TRANSMEM 105 125 POTENTIAL.  
 FT TRANSMEM 148 168 POTENTIAL.  
 FT TRANSMEM 175 195 POTENTIAL.  
 FT TRANSMEM 198 218 POTENTIAL.  
 FT TRANSMEM 225 245 POTENTIAL.  
 FT TRANSMEM 259 279 POTENTIAL.  
 FT TRANSMEM 295 315 POTENTIAL.  
 FT TRANSMEM 347 367 POTENTIAL.  
 FT TRANSMEM 372 392 POTENTIAL.  
 SQ SEQUENCE 404 AA; 42257 MW; 57D7D2807034A426 CRC64;  
  
 Query Match 26.0%; Score 47; DB 1; Length 404;  
 Best Local Similarity 43.5%; Pred. No. 18;  
 Matches 10; Conservative 1; Mismatches 12; Indels 0; Gaps 0;  
  
 OY 10 IGITELGGCHGSEPCNIHRGKPF 32  
 | : | | | | | | | | | |  
 Db 233 ISVVLGAGCLGFLPHNFRKIP 255  
  
 RESULT 11  
 BHMT\_HUMAN STANDARD; PRT; 406 AA.  
 ID BHMT\_HUMAN  
 AC Q93088;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE BETALINE--HOMOCYSTEINE S-METHYLTRANSFERASE (EC 2.1.1.5).  
 GN BHMT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=96394355; PubMed=8798461;

RC STRAIN=129; TISSUE=Liver;  
RA Sowden M.P., Smith H.C.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: INVOLVED IN THE REGULATION OF HOMOCYSTEINE METABOLISM.  
CC CONVERTS BETANINE AND HOMOCYSTEINE TO DIMETHYLGLYCINE AND  
CC METHIONINE, RESPECTIVELY. THIS REACTION IS ALSO REQUIRED FOR THE  
CC IRREVERSIBLE OXIDATION OF CHOLINE.  
CC CATALYTIC ACTIVITY: TRIMETHYLAMMONIOACETATE + L-HOMOCYSTEINE =  
CC DIMETHYLGLYCINE + L-METHIONINE.  
CC COFACTOR: ZINC (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF033381; AAB87501.1; -  
DR MGD; MGI:1339972; Bmmt.  
KW Transferase; Methyltransferase; Zinc.  
SQ SEQUENCE 407 AA; 45020 MW; 1C600BE9CC44EE32 CRC64;  
-----  
Query Match 26.0%; Score 47; DB 1; Length 407;  
Best Local Similarity 47.1%; Pred. No. 18;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
-----  
QY 10 IGITELGGCHGSEPCNI 26  
: : ||| | :  
Db 291 LGVRYIGCGGEPYHI 307  
-----  
RESULT 13  
BHMT\_RAT  
ID ID BHMT\_RAT STANDARD; PRT; 407 AA.  
AC AC09171;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DT BETANINE--HOMOCYSTEINE S-METHYLTRANSFERASE (EC 2.1.1.5).  
GN BHMT.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RS SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;  
RX MEDLINE=96354796; PubMed=8753772;  
RA Forestier M., Reichen J., Solioz M.;  
RT "Application of mRNA differential display to liver cirrhosis: reduced  
RT fetuin expression in biliary cirrhosis in the rat.";  
RL Biochem. Biophys. Res. Commun. 225:377-383(1996).  
RN [2]  
RS SEQUENCE FROM N.A.  
RC Sowden M.P., Sparks J.D., Sparks C.E., Smith H.C.;  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: INVOLVED IN THE REGULATION OF HOMOCYSTEINE METABOLISM.  
CC CONVERTS BETANINE AND HOMOCYSTEINE TO DIMETHYLGLYCINE AND  
CC METHIONINE, RESPECTIVELY. THIS REACTION IS ALSO REQUIRED FOR THE  
CC IRREVERSIBLE OXIDATION OF CHOLINE.  
CC CATALYTIC ACTIVITY: TRIMETHYLAMMONIOACETATE + L-HOMOCYSTEINE =  
CC DIMETHYLGLYCINE + L-METHIONINE.  
CC COFACTOR: ZINC (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U96133; AAB53763.1; -;  
 DR EMBL; AF038870; AAB95481.1; -;  
 KW Transferrase; Methyltransferase; Zinc.  
 SQ SEQUENCE 407 AA; 44976 MW; 36E1D04A8E425887 CRC64;

Query Match 26.0%; Score 47; DB 1; Length 407;  
 Best Local Similarity 47.1%; Pred. No. 18;  
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Oy 10 IGITELGGCHGSEPCNI 26  
 Db 291 LGVRYIGCGGPEPYHI 307

## RESULT 14

ID ALG2\_TYRPU STANDARD; PRT; 141 AA.  
 AC O02380;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE GROUP 2 ALLERGEN PRECURSOR.  
 OS Tyrophagus putrescentiae (Dust mite).  
 CC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
 CC Acariformes; Sarcopitiformes; Astigmata; Acaroidea; Acaridae;  
 CC Tyrophagus.  
 OX NCBI\_TaxID=59818;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Eriksson T., Johansson E., Whitley P., Schmidt M., Elsayed S.,  
 RA van Hage-Hamsten M.;  
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE EI FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; Y12690; CAA73221.1; -;  
 DR HSSP; Q00855; IAHM.  
 KW Allergen; Signal.  
 FT SIGNAL 1 15 POTENTIAL.  
 FT CHAIN 16 141 GROUP 2 ALLERGEN.  
 FT DISULFID 23 132 BY SIMILARITY.  
 FT DISULFID 36 41 BY SIMILARITY.  
 FT DISULFID 87 92 BY SIMILARITY.  
 FT CARBOHYD 103 103 N-LINKED (GLCNAC... ) (POTENTIAL).  
 SQ SEQUENCE 141 AA; 14851 MW; 38EF9520010A04C1 CRC64;

Query Match 25.7%; Score 46.5; DB 1; Length 141;  
 Best Local Similarity 32.3%; Pred. No. 7.9;  
 Matches 10; Conservative 5; Mismatches 15; Indels 1; Gaps 1;

Oy 1 QYIKANSFEIGTELGGCHGSEPCNIHRKPK 31  
 Db 19 KFTDGGKKEIASVAVDGCCEG-DLCVIHKSKP 48

## RESULT 15

METX\_HAEIN.  
 ID METX\_HAEIN STANDARD; PRT; 357 AA.  
 AC P45131;

DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE HOMOSERINE O-ACETYLTRANSFERASE (EC 2.3.1.31) (HOMOSERINE O-TRANS-  
 DE ACETYLASE) (HOMOSERINE TRANSACETYLASE) (HTA).  
 GN METX OR MET2 OR HI1263.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RD / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus  
 RT influenzae Rd.";  
 RL Science 269:496-512(1995).  
 RN [2]  
 CH CHARACTERIZATION.  
 RP MEDLINE=20374530; PubMed=10913262;  
 RA Born T.L., Franklin M., Blanchard J.S.;  
 RT "Enzyme-catalyzed acylation of homoserine: mechanistic  
 RT characterization of the Haemophilus influenzae met2-encoded  
 RT homoserine transacetylase.";  
 RL Biochemistry 39:8556-8564(2000).  
 CC -1- CATALYTIC ACTIVITY: ACETYL-COA + L-HOMOSERINE -> COA + O-ACETYL-L-  
 CC HOMOSERINE.  
 CC -1- PATHWAY: BIOSYNTHESIS OF METHIONINE; HTA VARIANT; FIRST STEP.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- MASS SPECTROMETRY: MW=39859; METHOD=ELECTROSPRAY.  
 CC -1- SIMILARITY: BELONGS TO THE AB HYDROLASE FAMILY; HTA SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U32806; AAC22916.1; -;  
 DR TIGR; HI1263; -;  
 DR InterPro; IPR000073; -;  
 DR Pfam; PF00561; abhydrolase; 1.  
 KW Methionine biosynthesis; Transferase; Acyltransferase.  
 FT INIT\_MET 0 0  
 FT ACT\_SITE 142 142 POTENTIAL.  
 FT ACT\_SITE 336 336 POTENTIAL.  
 SQ SEQUENCE 357 AA; 39859 MW; 99E6947EA7A8EC79 CRC64;

Query Match 25.7%; Score 46.5; DB 1; Length 357;  
 Best Local Similarity 40.7%; Pred. No. 19;  
 Matches 11; Conservative 2; Mismatches 11; Indels 3; Gaps 1;

Oy 9 FIGITELGGCHGS---EPCNIHRKPKF 32  
 Db 81 FISSVLGGCKGTGTPSSINPQTGPKY 107

Search completed: April 14, 2001, 10:17:46  
 Job time: 348 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2001, 10:09:35 ; Search time 41.8 Seconds  
(without alignments)  
14.247 Million cell updates/sec

Title: US-09-362-731-1

Perfect score: 165

Sequence: 1 QYIKANSKFIGITELGGHEIKKVLVPGCHGS 31

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

1: /cgn2.6/ptodata/2/iaa/5A\_COMB.pep.\*

2: /cgn2.6/ptodata/2/iaa/5B\_COMB.pep.\*

3: /cgn2.6/ptodata/2/iaa/6A\_COMB.pep.\*

4: /cgn2.6/ptodata/2/iaa/6B\_COMB.pep.\*

5: /cgn2.6/ptodata/2/iaa/PCTUS\_COMB.pep.\*

6: /cgn2.6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	92	55.8	47	1	US-08-446-692-35
2	92	55.8	47	2	US-08-488-351A-35
3	86	52.1	37	1	US-08-446-692-57
4	86	52.1	37	1	US-08-446-692-63
5	86	52.1	37	2	US-08-488-351A-57
6	86	52.1	37	2	US-08-488-351A-63
7	84	50.9	25	2	US-08-482-142-169
8	84	50.9	25	2	US-08-478-572-169
9	84	50.9	25	5	PCT-US95-04481-31
10	84	50.9	26	2	US-08-482-142-50
11	84	50.9	26	2	US-08-482-142-170
12	84	50.9	26	2	US-08-478-572-50
13	84	50.9	26	2	US-08-478-572-170
14	84	50.9	27	2	US-08-482-142-171
15	84	50.9	27	2	US-08-478-572-171
16	84	50.9	35	3	US-08-460-040-7
17	84	50.9	129	1	US-07-945-288-12
18	84	50.9	129	1	US-08-462-831-12
19	84	50.9	129	1	US-08-461-809-12
20	84	50.9	129	1	US-08-461-441-12
21	84	50.9	129	5	PCT-US93-08518-12
22	84	50.9	145	3	US-08-460-040-6
23	84	50.9	146	1	US-07-945-288-4
24	84	50.9	146	1	US-08-462-831-4
25	84	50.9	146	1	US-08-461-809-4
26	84	50.9	146	1	US-08-461-441-4
27	84	50.9	146	2	US-08-482-142-4

28	84	50.9	146	2	US-08-478-572-4	Sequence 4, Appl
29	84	50.9	146	5	PCT-US93-08518-4	Sequence 4, Appl
30	83	50.3	26	2	US-08-482-142-52	Sequence 52, Appl
31	83	50.3	26	2	US-08-478-572-52	Sequence 52, Appl
32	82	49.7	26	2	US-08-482-142-51	Sequence 51, Appl
33	82	49.7	26	2	US-08-478-572-51	Sequence 51, Appl
34	79	47.9	16	2	US-08-482-142-59	Sequence 59, Appl
35	79	47.9	16	2	US-08-478-572-59	Sequence 59, Appl
36	79	47.9	25	2	US-08-482-142-42	Sequence 42, Appl
37	79	47.9	25	2	US-08-478-572-42	Sequence 42, Appl
38	79	47.9	25	5	PCT-US95-04481-19	Sequence 19, Appl
39	75	45.5	31	5	PCT-US93-11703-63	Sequence 63, Appl
40	75	45.5	61	2	US-08-482-142-162	Sequence 162, App
41	75	45.5	61	2	US-08-478-572-162	Sequence 162, App
42	75	45.5	84	2	US-08-482-142-161	Sequence 161, App
43	75	45.5	84	2	US-08-478-572-161	Sequence 161, App
44	75	45.5	129	2	US-08-482-142-157	Sequence 157, App
45	75	45.5	129	2	US-08-482-142-159	Sequence 159, App

#### ALIGNMENTS

RESULT 1  
US-08-446-692-35  
; Sequence 35, Application US/08446692  
; Patent No. 5759551  
; GENERAL INFORMATION:  
; APPLICANT: Ladd, Anna  
; APPLICANT: Wang, Chang Yi  
; APPLICANT: Zamb, Timothy  
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
; TITLE OF INVENTION: and synthetic universal immune stimulants for vaccines  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Maria C.H. Lin  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10154-0053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/446,692  
; FILING DATE: 7-JUN-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maria C.H. Lin  
; REGISTRATION NUMBER: 29,323  
; REFERENCE/DOCKET NUMBER: 1151-4146 US2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)415-8745  
; TELEFAX: (516)751-6849  
; INFORMATION FOR SEQ ID NO: 35:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 47 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-446-692-35

Query Match 55.8%; Score 92; DB 1; Length 47;  
Best Local Similarity 74.1%; Pred. No. 3.5e-07;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITELGGHEIKKVLVPG 27  
|||||

Db 21 QYIKANSKFIGITELGGHEIWSYGLRPG 47  
|||||

## RESULT 2

US-08-488-351A-35  
; Sequence 35, Application US/08488351A  
; Patent No. 5843446  
; GENERAL INFORMATION:  
; APPLICANT: Ladd, Anna  
; APPLICANT: Wang, Chang Yi  
; APPLICANT: Zamb, Timothy  
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Maria C.H. Lin  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10154-0053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/488,351A  
; FILING DATE: 7-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/229,275  
; FILING DATE: 14-APR-1994  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/057,166  
; FILING DATE: 27-APR-1992  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maria C.H. Lin  
; REGISTRATION NUMBER: 29,323  
; REFERENCE/DOCKET NUMBER: 1151-4146 US2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)415-8745  
; TELEFAX: (516)751-6849  
; INFORMATION FOR SEQ ID NO: 35:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 47 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-488-351A-35

Query Match 55.88; Score 92; DB 2; Length 47;  
Best Local Similarity 74.1%; Pred. No. 3.5e-07;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITELGGHEIKKVLVPG 27  
Db 21 QYIKANSKFIGITELGGHWSYGLRPG 47

## RESULT 3

US-08-446-692-57  
; Sequence 57, Application US/08446692  
; Patent No. 5759551  
; GENERAL INFORMATION:  
; APPLICANT: Ladd, Anna  
; APPLICANT: Wang, Chang Yi

; APPLICANT: Zamb, Timothy  
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Maria C.H. Lin  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10154-0053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/446,692  
; FILING DATE: 7-JUN-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maria C.H. Lin  
; REGISTRATION NUMBER: 29,323  
; REFERENCE/DOCKET NUMBER: 1151-4146 US2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)415-8745  
; TELEFAX: (516)751-6849  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 37 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-446-692-57

Query Match 52.1%; Score 86; DB 1; Length 37;  
Best Local Similarity 100.0%; Pred. No. 2.2e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITELGG 17  
Db 21 QYIKANSKFIGITELGG 37

## RESULT 4

US-08-446-692-63  
; Sequence 63, Application US/08446692  
; Patent No. 5759551  
; GENERAL INFORMATION:  
; APPLICANT: Ladd, Anna  
; APPLICANT: Wang, Chang Yi  
; APPLICANT: Zamb, Timothy  
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Maria C.H. Lin  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10154-0053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/446,692  
; FILING DATE: 7-JUN-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:



NAME: Maria C.H. Lin  
REGISTRATION NUMBER: 29,323  
REFERENCE/DOCKET NUMBER: 1151-4146 US2  
TELEPHONE: (212)415-8745  
TELEFAX: (516)751-6849  
INFORMATION FOR SEQ ID NO: 63:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 37 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-446-692-63

Query Match 52.1%; Score 86; DB 1; Length 37;  
Best Local Similarity 100.0%; Pred. No. 2.2e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITELGG 17  
|||||  
DB 5 QYIKANSKFIGITELGG 21

RESULT 5  
US-08-488-351A-57  
; Sequence 57, Application US/08488351A  
; Patent No. 5843446  
; GENERAL INFORMATION:  
; APPLICANT: Ladd, Anna  
; APPLICANT: Wang, Chang Yi  
; APPLICANT: Zamb, Timothy  
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Maria C.H. Lin  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10154-0053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE: 7-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION NUMBER: US/08/488,351A  
; APPLICATION NUMBER: US 08/446,692  
; FILING DATE: 7-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/229,275  
; FILING DATE: 14-APR-1994  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/057,166  
; FILING DATE: 27-APR-1992  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/057,166  
; FILING DATE: 27-APR-1992  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maria C.H. Lin  
; REGISTRATION NUMBER: 29,323  
; REFERENCE/DOCKET NUMBER: 1151-4146 US2  
; TELEPHONE: (212)415-8745  
; TELEFAX: (516)751-6849  
; INFORMATION FOR SEQ ID NO: 57:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 37 amino acids

NAME: Maria C.H. Lin  
REGISTRATION NUMBER: 29,323  
REFERENCE/DOCKET NUMBER: 1151-4146 US2  
TELEPHONE: (212)415-8745  
TELEFAX: (516)751-6849  
INFORMATION FOR SEQ ID NO: 63:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 37 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-488-351A-57

Query Match 52.1%; Score 86; DB 2; Length 37;  
Best Local Similarity 100.0%; Pred. No. 2.2e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITELGG 17  
|||||  
DB 21 QYIKANSKFIGITELGG 37

RESULT 6  
US-08-488-351A-63  
; Sequence 63, Application US/08488351A  
; Patent No. 5843446  
; GENERAL INFORMATION:  
; APPLICANT: Ladd, Anna  
; APPLICANT: Wang, Chang Yi  
; APPLICANT: Zamb, Timothy  
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Maria C.H. Lin  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10154-0053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/488,351A  
; FILING DATE: 7-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/446,692  
; FILING DATE: 7-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/229,275  
; FILING DATE: 14-APR-1994  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/057,166  
; FILING DATE: 27-APR-1992  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maria C.H. Lin  
; REGISTRATION NUMBER: 29,323  
; REFERENCE/DOCKET NUMBER: 1151-4146 US2  
; TELEPHONE: (212)415-8745  
; TELEFAX: (516)751-6849  
; INFORMATION FOR SEQ ID NO: 63:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 37 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-488-351A-63

Query Match 52.1%; Score 86; DB 2; Length 37;  
Best Local Similarity 100.0%; Pred. No. 2.2e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITELGG 17  
Db 5 QYIKANSKFIGITELGG 21

## RESULT 7

US-08-482-142-169  
Sequence 169, Application US/08482142  
Patent No. 5820862

## GENERAL INFORMATION:

APPLICANT: Garman, Richard  
APPLICANT: Greenstein, Julia  
APPLICANT: Kuo, Mei-chang  
APPLICANT: Rogers, Bruce  
APPLICANT: Franzen, Henry  
APPLICANT: Chen, Xian  
APPLICANT: Evans, Sean  
APPLICANT: Shaked, Ze'ev  
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
NUMBER OF SEQUENCES: 207  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
STREET: 610 LINCOLN STREET  
CITY: WALTHAM  
STATE: MA  
COUNTRY: USA  
ZIP: 02154

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,142  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/445,307  
FILING DATE: 07 June 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: CRAIG, ANNE I.  
REGISTRATION NUMBER: 32,976  
REFERENCE/DOCKET NUMBER: 017.605  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 466-6000  
TELEFAX: (617) 466-6040  
INFORMATION FOR SEQ ID NO: 169:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal

US-08-482-142-169

Query Match 50.9%; Score 84; DB 2; Length 25;  
Best Local Similarity 63.6%; Pred. No. 2.9e-06;  
Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 10 IGITELGGHEIKKVLVPGCHGS 31  
Db 3 VDVKDCANHEIKKVLVPGCHGS 24

## RESULT 8

US-08-478-572-169  
Sequence 169, Application US/08478572  
Patent No. 5968526

## GENERAL INFORMATION:

APPLICANT: Garman, Richard

APPLICANT: Greenstein, Julia  
APPLICANT: Kuo, Mei-chang  
APPLICANT: Rogers, Bruce  
APPLICANT: Franzen, Henry  
APPLICANT: Chen, Xian  
APPLICANT: Evans, Sean  
APPLICANT: Shaked, Ze'ev  
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
NUMBER OF SEQUENCES: 207  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
STREET: 610 LINCOLN STREET  
CITY: WALTHAM  
STATE: MA  
COUNTRY: USA  
ZIP: 02154

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/478,572  
FILING DATE: 07-June-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/445,307  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: CRAIG, ANNE I.  
REGISTRATION NUMBER: 32,976  
REFERENCE/DOCKET NUMBER: 017.605  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 466-6000  
TELEFAX: (617) 466-6040  
INFORMATION FOR SEQ ID NO: 169:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-478-572-169

Query Match 50.9%; Score 84; DB 2; Length 25;  
Best Local Similarity 63.6%; Pred. No. 2.9e-06;  
Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 10 IGITELGGHEIKKVLVPGCHGS 31  
Db 3 VDVKDCANHEIKKVLVPGCHGS 24

## RESULT 9

PCT-US95-04481-31  
Sequence 31, Application PC/TUS9504481  
GENERAL INFORMATION:

APPLICANT:  
TITLE OF INVENTION: Pharmaceutical Peptide Formulations For Treatment of Dust M  
NUMBER OF SEQUENCES: 54  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04481  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/227,772

;; FILING DATE: April 14, 1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Vanstone, Darlene A.  
;; REGISTRATION NUMBER: 35,279  
;; REFERENCE/DOCKET NUMBER: 017.5 PCT  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 466-6000  
;; TELEFAX: (617) 466-6040  
;; INFORMATION FOR SEQ ID NO: 31:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 25 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
PCT-US95-04481-31

Query Match 50.9%; Score 84; DB 5; Length 25;  
Best Local Similarity 63.6%; Pred. No. 2.9e-06;  
Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 10 IGITELGGHEIKKVLVPGCHGS 31  
Db 3 VDKDCANHEIKKVLVPGCHGS 24

RESULT 10  
US-08-482-142-50  
; Sequence 50, Application US/08482142  
; Patent No. 5820862  
; GENERAL INFORMATION:  
; APPLICANT: Garman, Richard  
; APPLICANT: Greenstein, Julia  
; APPLICANT: Kuo, Mei-chang  
; APPLICANT: Rogers, Bruce  
; APPLICANT: Franzen, Henry  
; APPLICANT: Chen, Xian  
; APPLICANT: Evans, Sean  
; APPLICANT: Shaked, Ze'ev  
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
; STREET: 610 LINCOLN STREET  
; CITY: WALTHAM  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII TEXT  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/482,142  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/445,307  
; FILING DATE: 07 June 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CRAIG, ANNE I.  
; REGISTRATION NUMBER: 32,976  
; REFERENCE/DOCKET NUMBER: 017.6US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 466-6000  
; TELEFAX: (617) 466-6040  
; INFORMATION FOR SEQ ID NO: 50:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 26 amino acids  
; TYPE: amino acid

;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FRAGMENT TYPE: N-terminal  
US-08-482-142-50  
  
Query Match 50.9%; Score 84; DB 2; Length 26;  
Best Local Similarity 63.6%; Pred. No. 3e-06;  
Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
  
QY 10 IGITELGGHEIKKVLVPGCHGS 31  
Db 3 VDKDCANHEIKKVLVPGCHGS 24

RESULT 11  
US-08-482-142-170  
; Sequence 170, Application US/08482142  
; Patent No. 5820862  
; GENERAL INFORMATION:  
; APPLICANT: Garman, Richard  
; APPLICANT: Greenstein, Julia  
; APPLICANT: Kuo, Mei-chang  
; APPLICANT: Rogers, Bruce  
; APPLICANT: Franzen, Henry  
; APPLICANT: Chen, Xian  
; APPLICANT: Evans, Sean  
; APPLICANT: Shaked, Ze'ev  
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
; STREET: 610 LINCOLN STREET  
; CITY: WALTHAM  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII TEXT  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/482,142  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/445,307  
; FILING DATE: 07 June 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CRAIG, ANNE I.  
; REGISTRATION NUMBER: 32,976  
; REFERENCE/DOCKET NUMBER: 017.6US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 466-6000  
; TELEFAX: (617) 466-6040  
; INFORMATION FOR SEQ ID NO: 170:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 26 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
US-08-482-142-170

Query Match 50.9%; Score 84; DB 2; Length 26;  
Best Local Similarity 63.6%; Pred. No. 3e-06;  
Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 10 IGITELGGHEIKKVLVPGCHGS 31  
; : : : |||||

Db 3 VDVKDCANHEIKKVLVPGCHGS 24

## RESULT 12

US-08-478-572-50  
; Sequence 50, Application US/08478572  
; Patent No. 5968526

## ; GENERAL INFORMATION:

; APPLICANT: Garman, Richard  
; APPLICANT: Greenstein, Julia  
; APPLICANT: Kuo, Mei-chang  
; APPLICANT: Rogers, Bruce  
; APPLICANT: Franzen, Henry  
; APPLICANT: Chen, Xian  
; APPLICANT: Evans, Sean  
; APPLICANT: Shaked, Ze'ev  
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
; NUMBER OF SEQUENCES: 207

## ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
; STREET: 610 LINCOLN STREET  
; CITY: WALTHAM  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII TEXT  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/478,572  
; FILING DATE: 07-June-1995

## ; CLASSIFICATION:

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/445,307  
; FILING DATE:

## ; ATTORNEY/AGENT INFORMATION:

; NAME: CRAIG, ANNE I.  
; REGISTRATION NUMBER: 32,976  
; REFERENCE/DOCKET NUMBER: 017.6US  
; TELEPHONE: (617) 466-6000  
; TELEFAX: (617) 466-6040  
; INFORMATION FOR SEQ ID NO: 50:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 26 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; FRAGMENT TYPE: N-terminal

US-08-478-572-50

Query Match 50.9%; Score 84; DB 2; Length 26;  
Best Local Similarity 63.6%; Pred. No. 3e-06;  
Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 10 IGITELGGHEIKKVLVPGCHGS 31

: : : |||||

Db 3 VDVKDCANHEIKKVLVPGCHGS 24

## RESULT 13

US-08-478-572-170  
; Sequence 170, Application US/08478572  
; Patent No. 5968526

## ; GENERAL INFORMATION:

; APPLICANT: Garman, Richard  
; APPLICANT: Greenstein, Julia  
; APPLICANT: Kuo, Mei-chang  
; APPLICANT: Rogers, Bruce

; APPLICANT: Franzen, Henry  
; APPLICANT: Chen, Xian  
; APPLICANT: Evans, Sean  
; APPLICANT: Shaked, Ze'ev  
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
; STREET: 610 LINCOLN STREET  
; CITY: WALTHAM  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII TEXT  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/478,572  
; FILING DATE: 07-June-1995

## ; CLASSIFICATION:

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/445,307  
; FILING DATE:

## ; ATTORNEY/AGENT INFORMATION:

; NAME: CRAIG, ANNE I.  
; REGISTRATION NUMBER: 32,976  
; REFERENCE/DOCKET NUMBER: 017.6US  
; TELEPHONE: (617) 466-6000  
; TELEFAX: (617) 466-6040  
; INFORMATION FOR SEQ ID NO: 170:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 26 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; FRAGMENT TYPE: internal

US-08-478-572-170

Query Match 50.9%; Score 84; DB 2; Length 26;  
Best Local Similarity 63.6%; Pred. No. 3e-06;  
Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 10 IGITELGGHEIKKVLVPGCHGS 31

: : : |||||

Db 3 VDVKDCANHEIKKVLVPGCHGS 24

## RESULT 14

US-08-482-142-171  
; Sequence 171, Application US/08482142  
; Patent No. 5820862

## ; GENERAL INFORMATION:

; APPLICANT: Garman, Richard  
; APPLICANT: Greenstein, Julia  
; APPLICANT: Kuo, Mei-chang  
; APPLICANT: Rogers, Bruce  
; APPLICANT: Franzen, Henry  
; APPLICANT: Chen, Xian  
; APPLICANT: Evans, Sean  
; APPLICANT: Shaked, Ze'ev

; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
; NUMBER OF SEQUENCES: 207

## ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
; STREET: 610 LINCOLN STREET  
; CITY: WALTHAM  
; STATE: MA

;; COUNTRY: USA  
;; ZIP: 02154  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: ASCII TEXT  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/482,142  
;; FILING DATE: 07-JUN-1995  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/445,307  
;; FILING DATE: 07 June 1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: CRAIG, ANNE I.  
;; REGISTRATION NUMBER: 32,976  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 466-6000  
;; TELEFAX: (617) 466-6040  
;; INFORMATION FOR SEQ ID NO: 171:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 27 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FRAGMENT TYPE: internal  
;; US-08-482-142-171

Query Match 50.9%; Score 84; DB 2; Length 27;  
Best Local Similarity 63.6%; Pred. No. 3.1e-06;  
Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 10 IGITELGGHEIKKVLVPGCHGS 31  
: : : |||||  
Db 3 VDVKDCANHEIKKVLVPGCHGS 24

## RESULT 15

US-08-478-572-171  
; Sequence 171, Application US/08478572  
; Patent No. 5968526  
; GENERAL INFORMATION:  
; APPLICANT: Garman, Richard  
; APPLICANT: Greenstein, Julia  
; APPLICANT: Kuo, Mei-chang  
; APPLICANT: Rogers, Bruce  
; APPLICANT: Franzen, Henry  
; APPLICANT: Chen, Xian  
; APPLICANT: Evans, Sean  
; APPLICANT: Shaked, Ze'ev  
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
; FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
; STREET: 610 LINCOLN STREET  
; CITY: WALTHAM  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII TEXT  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/478,572  
; FILING DATE: 07-June-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: 08/445,307  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: CRAIG, ANNE I.  
;; REGISTRATION NUMBER: 32,976  
;; REFERENCE/DOCKET NUMBER: 017.6US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 466-6000  
;; TELEFAX: (617) 466-6040  
;; INFORMATION FOR SEQ ID NO: 171:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 27 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FRAGMENT TYPE: internal  
;; US-08-478-572-171

Query Match 50.9%; Score 84; DB 2; Length 27;  
Best Local Similarity 63.6%; Pred. No. 3.1e-06;  
Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 10 IGITELGGHEIKKVLVPGCHGS 31  
: : : |||||  
Db 3 VDVKDCANHEIKKVLVPGCHGS 24

Search completed: April 14, 2001, 10:16:54  
Job time: 439 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2001, 10:16:56 ; Search time 41.8 Seconds  
(without alignments)  
14.707 Million cell updates/sec

Title: US-09-362-731-5

Perfect score: 181

Sequence: 1 QYIKANSKFIGITELGGCHGSEPCNIHRGKPF 32

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2.6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2.6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2.6/ptodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgn2.6/ptodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgn2.6/ptodata/2/iaa/PCTUS\_COMB.pep.\*
- 6: /cgn2.6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92.5	51.1	92	2	US-08-482-142-160
2	92.5	51.1	92	2	US-08-478-572-160
3	92	50.8	25	2	US-08-482-142-42
4	92	50.8	25	2	US-08-478-572-42
5	92	50.8	25	5	PCT-US95-04481-19
6	92	50.8	35	3	US-08-460-040-7
7	92	50.8	47	1	US-08-446-692-35
8	92	50.8	47	2	US-08-488-351A-35
9	92	50.8	61	2	US-08-482-142-162
10	92	50.8	61	2	US-08-478-572-162
11	92	50.8	84	2	US-08-482-142-161
12	92	50.8	84	2	US-08-478-572-161
13	92	50.8	129	1	US-07-945-288-12
14	92	50.8	129	1	US-08-462-831-12
15	92	50.8	129	1	US-08-461-809-12
16	92	50.8	129	1	US-08-461-441-12
17	92	50.8	129	2	US-08-482-142-157
18	92	50.8	129	2	US-08-482-142-159
19	92	50.8	129	2	US-08-478-572-157
20	92	50.8	129	2	US-08-478-572-159
21	92	50.8	129	5	PCT-US93-08518-12
22	92	50.8	130	2	US-08-482-142-158
23	92	50.8	130	2	US-08-478-572-158
24	92	50.8	145	3	US-08-460-040-6
25	92	50.8	146	1	US-07-945-288-4
26	92	50.8	146	1	US-08-462-831-4
27	92	50.8	146	1	US-08-461-809-4

28	92	50.8	146	1	US-08-461-441-4	Sequence 4, Appli
29	92	50.8	146	2	US-08-482-142-4	Sequence 4, Appli
30	92	50.8	146	2	US-08-478-572-4	Sequence 4, Appli
31	92	50.8	146	5	PCT-US93-08518-4	Sequence 4, Appli
32	90	49.7	129	4	US-08-930-264-18	Sequence 18, Appli
33	89	49.2	25	2	US-08-482-142-104	Sequence 104, App
34	89	49.2	25	2	US-08-478-572-104	Sequence 104, App
35	89	49.2	129	1	US-07-945-288-8	Sequence 8, Appli
36	89	49.2	129	1	US-07-945-288-13	Sequence 13, Appli
37	89	49.2	129	1	US-08-462-831-8	Sequence 8, Appli
38	89	49.2	129	1	US-08-462-831-13	Sequence 13, Appli
39	89	49.2	129	1	US-08-461-809-8	Sequence 8, Appli
40	89	49.2	129	1	US-08-461-809-13	Sequence 13, Appli
41	89	49.2	129	1	US-08-461-441-8	Sequence 8, Appli
42	89	49.2	129	1	US-08-461-441-13	Sequence 13, Appli
43	89	49.2	129	2	US-08-482-142-8	Sequence 8, Appli
44	89	49.2	129	2	US-08-478-572-8	Sequence 8, Appli
45	89	49.2	129	5	PCT-US93-08518-13	Sequence 13, Appli

#### ALIGNMENTS

RESULT 1  
US-08-482-142-160  
: Sequence 160, Application US/08482142  
: Patent No. 5820862  
: GENERAL INFORMATION:  
: APPLICANT: Garman, Richard  
: APPLICANT: Greenstein, Julia  
: APPLICANT: Kuo, Mei-chang  
: APPLICANT: Rogers, Bruce  
: APPLICANT: Franzen, Henry  
: APPLICANT: Chen, Xian  
: APPLICANT: Evans, Sean  
: APPLICANT: Shaked, Ze'ev  
: TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
: NUMBER OF SEQUENCES: 207  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
: STREET: 610 LINCOLN STREET  
: CITY: WALTHAM  
: STATE: MA  
: COUNTRY: USA  
: ZIP: 02154  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: ASCII TEXT  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/482,142  
: FILING DATE: 07-JUN-1995  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US/08/445,307  
: FILING DATE: 07 June 1995  
: ATTORNEY/AGENT INFORMATION:  
: NAME: CRAIG, ANNE I.  
: REGISTRATION NUMBER: 32,976  
: REFERENCE/DOCKET NUMBER: 017,605  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (617) 466-6000  
: TELEFAX: (617) 466-6040  
: INFORMATION FOR SEQ ID NO: 160:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 92 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear  
: MOLECULE TYPE: peptide  
: FRAGMENT TYPE: internal  
US-08-482-142-160

Query Match 51.1%; Score 92.5; DB 2; Length 92;  
Best Local Similarity 67.9%; Pred. No. 5.8e-06;  
Matches 19; Conservative 2; Mismatches 6; Indels 1; Gaps 1;

Qy 5 ANSKFIGITELGGCHGSEPCNHRGKPF 32  
| | | : | | | | | | | | | | | | | |  
Db 9 ANSE-IKKVMVPGCHGSEPCIIHRGKPF 35

## RESULT 2

US-08-478-572-160  
; Sequence 160, Application US/08478572  
; Patent No. 5968526  
; GENERAL INFORMATION:  
; APPLICANT: Garman, Richard  
; APPLICANT: Greenstein, Julia  
; APPLICANT: Kuo, Mei-chang  
; APPLICANT: Rogers, Bruce  
; APPLICANT: Franzen, Henry  
; APPLICANT: Chen, Xian  
; APPLICANT: Evans, Sean  
; APPLICANT: Shaked, Ze'ev  
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
; STREET: 610 LINCOLN STREET  
; CITY: WALTHAM  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/478.572  
FILING DATE: 07-June-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/445,307  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: CRAIG, ANNE I.  
REGISTRATION NUMBER: 32,976  
REFERENCE/DOCKET NUMBER: 017.6US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 466-6000  
TELEFAX: (617) 466-6040  
INFORMATION FOR SEQ ID NO: 160:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 92 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-478-572-160

Query Match 51.1%; Score 92.5; DB 2; Length 92;  
Best Local Similarity 67.9%; Pred. No. 5.8e-06;  
Matches 19; Conservative 2; Mismatches 6; Indels 1; Gaps 1;

Qy 5 ANSKFIGITELGGCHGSEPCNHRGKPF 32  
| | | : | | | | | | | | | | | | | |  
Db 3 ANSE-IKKVMVPGCHGSEPCIIHRGKPF 35

## RESULT 3

US-08-482-142-42  
; Sequence 42, Application US/08482142  
; Patent No. 5820862  
; GENERAL INFORMATION:  
; APPLICANT: Garman, Richard  
; APPLICANT: Greenstein, Julia  
; APPLICANT: Kuo, Mei-chang  
; APPLICANT: Rogers, Bruce  
; APPLICANT: Franzen, Henry  
; APPLICANT: Chen, Xian  
; APPLICANT: Evans, Sean  
; APPLICANT: Shaked, Ze'ev  
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
; STREET: 610 LINCOLN STREET  
; CITY: WALTHAM  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,142  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/445,307  
FILING DATE: 07 June 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: CRAIG, ANNE I.  
REGISTRATION NUMBER: 32,976  
REFERENCE/DOCKET NUMBER: 017.6US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 466-6000  
TELEFAX: (617) 466-6040  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-482-142-42

Query Match 50.8%; Score 92; DB 2; Length 25;  
Best Local Similarity 93.8%; Pred. No. 1.8e-06;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 17 GCHGSEPCNHRGKPF 32  
| | | | | | | | | | | | | | | | | | | |  
Db 10 GCHGSEPCIIHRGKPF 25

## RESULT 4

US-08-478-572-42  
; Sequence 42, Application US/08478572  
; Patent No. 5968526  
; GENERAL INFORMATION:  
; APPLICANT: Garman, Richard  
; APPLICANT: Greenstein, Julia  
; APPLICANT: Kuo, Mei-chang  
; APPLICANT: Rogers, Bruce  
; APPLICANT: Franzen, Henry  
; APPLICANT: Chen, Xian  
; APPLICANT: Evans, Sean  
; APPLICANT: Shaked, Ze'ev



;; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
;; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
;; NUMBER OF SEQUENCES: 207  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
;; STREET: 610 LINCOLN STREET  
;; CITY: WALTHAM  
;; STATE: MA  
;; COUNTRY: USA  
;; ZIP: 02154  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: ASCII TEXT  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/478,572  
;; FILING DATE: 07-June-1995  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/445,307  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: CRAIG, ANNE I.  
;; REGISTRATION NUMBER: 32,976  
;; REFERENCE/DOCKET NUMBER: 017.6US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 466-6000  
;; TELEFAX: (617) 466-6040  
;; INFORMATION FOR SEQ ID NO: 42:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 25 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FRAGMENT TYPE: N-terminal  
US-08-478-572-42

Query Match 50.8%; Score 92; DB 2; Length 25;  
Best Local Similarity 93.8%; Pred. No. 1.8e-06;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 17 GCHGSEPCNIHRGKPF 32  
||||||| |||||||  
Db 10 GCHGSEPCIIHRGAPF 25

RESULT 5  
PCT-US95-04481-19  
;; Sequence 19, Application PC/TUS9504481  
;; GENERAL INFORMATION:  
;; APPLICANT:  
;; TITLE OF INVENTION: Pharmaceutical Peptide Formulations For Treatment of Dust Mit  
;; NUMBER OF SEQUENCES: 54  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US95/04481  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/227,772  
;; FILING DATE: April 14, 1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Vanstone, Darlene A.  
;; REGISTRATION NUMBER: 35,279  
;; REFERENCE/DOCKET NUMBER: 017.5 PCT  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 466-6000

;; TELEFAX: (617) 466-6040  
;; INFORMATION FOR SEQ ID NO: 19:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 25 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FRAGMENT TYPE: internal  
PCT-US95-04481-19

Query Match 50.8%; Score 92; DB 5; Length 25;  
Best Local Similarity 93.8%; Pred. No. 1.8e-06;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 17 GCHGSEPCNIHRGKPF 32  
||||||| |||||||  
Db 10 GCHGSEPCIIHRGKPF 25

RESULT 6  
US-08-460-040-7  
;; Sequence 7, Application US/08460040  
;; Patent No. 6071522  
;; GENERAL INFORMATION:  
;; APPLICANT: Thomas, Wayne R.  
;; TITLE OF INVENTION: Cloning of Mite Allergens  
;; NUMBER OF SEQUENCES: 8  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: LAHIVE & COCKFIELD  
;; STREET: 60 State Street, suite 510  
;; CITY: Boston  
;; STATE: Massachusetts  
;; COUNTRY: USA  
;; ZIP: 02109-1875  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/460,040  
;; FILING DATE: 2-JUNE-95  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/162,722  
;; FILING DATE: 8-NOV-93  
;; APPLICATION NUMBER: 07/458,642  
;; FILING DATE: 13-FEB-90  
;; APPLICATION NUMBER: PCT/AU88/00195  
;; FILING DATE: 17-JUNE-88  
;; APPLICATION NUMBER: PI 2523/87  
;; FILING DATE: 18-JUNE-87  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Amy E. Mandragouras  
;; REGISTRATION NUMBER: 36,207  
;; REFERENCE/DOCKET NUMBER: IMI-021CN2  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617)227-7400  
;; TELEFAX: (617)227-5941  
;; INFORMATION FOR SEQ ID NO: 7:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 35 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FRAGMENT TYPE: internal  
US-08-460-040-7

Query Match 50.8%; Score 92; DB 3; Length 35;  
Best Local Similarity 93.8%; Pred. No. 2.5e-06;



COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,142  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/445,307  
FILING DATE: 07 June 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: CRAIG, ANNE I.  
REGISTRATION NUMBER: 32,976  
REFERENCE/DOCKET NUMBER: 017.6US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 466-6000  
TELEFAX: (617) 466-6040  
INFORMATION FOR SEQ ID NO: 162:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 61 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-482-142-162

Query Match 50.8%; Score 92; DB 2; Length 61;  
Best Local Similarity 93.8%; Pred. No. 4.5e-06;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 17 GCHGSEPCNIHRGKPF 32  
Db 20 GCHGSEPCIIHRGKPF 35

## RESULT 10

US-08-478-572-162  
Sequence 162, Application US/08478572  
Patent No. 5968526  
GENERAL INFORMATION:  
APPLICANT: Garman, Richard  
APPLICANT: Greenstein, Julia  
APPLICANT: Kuo, Mei-chang  
APPLICANT: Rogers, Bruce  
APPLICANT: Franzen, Henry  
APPLICANT: Chen, Xian  
APPLICANT: Evans, Sean  
APPLICANT: Shaked, Ze'ev  
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
NUMBER OF SEQUENCES: 207  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
STREET: 610 LINCOLN STREET  
CITY: WALTHAM  
STATE: MA  
COUNTRY: USA  
ZIP: 02154

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/478,572  
FILING DATE: 07-June-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/445,307  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: CRAIG, ANNE I.  
REGISTRATION NUMBER: 32,976  
REFERENCE/DOCKET NUMBER: 017.6US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 466-6000  
TELEFAX: (617) 466-6040  
INFORMATION FOR SEQ ID NO: 162:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 61 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-478-572-162

Query Match 50.8%; Score 92; DB 2; Length 61;  
Best Local Similarity 93.8%; Pred. No. 4.5e-06;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 17 GCHGSEPCNIHRGKPF 32  
Db 20 GCHGSEPCIIHRGKPF 35

## RESULT 11

US-08-482-142-161  
Sequence 161, Application US/08482142  
Patent No. 5820862  
GENERAL INFORMATION:  
APPLICANT: Garman, Richard  
APPLICANT: Greenstein, Julia  
APPLICANT: Kuo, Mei-chang  
APPLICANT: Rogers, Bruce  
APPLICANT: Franzen, Henry  
APPLICANT: Chen, Xian  
APPLICANT: Evans, Sean  
APPLICANT: Shaked, Ze'ev  
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
NUMBER OF SEQUENCES: 207  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
STREET: 610 LINCOLN STREET  
CITY: WALTHAM  
STATE: MA  
COUNTRY: USA  
ZIP: 02154

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,142  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/445,307  
FILING DATE: 07 June 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: CRAIG, ANNE I.  
REGISTRATION NUMBER: 32,976  
REFERENCE/DOCKET NUMBER: 017.6US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 466-6000  
TELEFAX: (617) 466-6040  
INFORMATION FOR SEQ ID NO: 161:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 84 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
US-08-482-142-161

Query Match 50.8%; Score 92; DB 2; Length 84;  
Best Local Similarity 93.8%; Pred. No. 6.2e-06;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 17 GCHGSEPCNIHRGKPF 32  
||||||| |||||||  
Db 20 GCHGSEPCIIHRGKPF 35

RESULT 12  
US-08-478-572-161  
; Sequence 161, Application US/08478572  
; Patent No. 5968526  
; GENERAL INFORMATION:  
; APPLICANT: Garman, Richard  
; APPLICANT: Greenstein, Julia  
; APPLICANT: Kuo, Mei-chang  
; APPLICANT: Rogers, Bruce  
; APPLICANT: Franzen, Henry  
; APPLICANT: Chen, Xian  
; APPLICANT: Evans, Sean  
; APPLICANT: Shaked, Ze'ev  
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
; STREET: 610 LINCOLN STREET  
; CITY: WALTHAM  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII TEXT  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/478,572  
; FILING DATE: 07-June-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/445,307  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CRAIG, ANNE I.  
; REGISTRATION NUMBER: 32,976  
; REFERENCE/DOCKET NUMBER: 017,605  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 466-6000  
; TELEFAX: (617) 466-6040  
; INFORMATION FOR SEQ ID NO: 161:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 84 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
US-08-478-572-161

Query Match 50.8%; Score 92; DB 2; Length 84;  
Best Local Similarity 93.8%; Pred. No. 6.2e-06;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 17 GCHGSEPCNIHRGKPF 32  
||||||| |||||||  
Db 20 GCHGSEPCIIHRGKPF 35

RESULT 13  
US-07-945-288-12  
; Sequence 12, Application US/07945288  
; Patent No. 5433948  
; GENERAL INFORMATION:  
; APPLICANT: Thomas, Wayne R.  
; APPLICANT: Chua, Kaw-Yan  
; TITLE OF INVENTION: CLONING AND SEQUENCING OF ALLERGENS FROM  
; TITLE OF INVENTION: DERMATOPHAGOIDES (HOUSE DUST MITES)  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 STATE STREET, SUITE 510  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII TEXT  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/945,288  
; FILING DATE: 19920910  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 580,655  
; FILING DATE: 11 SEPTEMBER 1990  
; APPLICATION NUMBER: 458,642  
; FILING DATE: 13 FEBRUARY 1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MANDRAGOURAS, AMY E.  
; REGISTRATION NUMBER: P36,207  
; REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 129 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 47  
; OTHER INFORMATION: /label-xaa is Thr or Ser  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 113  
; OTHER INFORMATION: /label-xaa is Asp or Asn  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 127  
; OTHER INFORMATION: /label-xaa is Ile or Leu  
US-07-945-288-12

Query Match 50.8%; Score 92; DB 1; Length 129;  
Best Local Similarity 93.8%; Pred. No. 9.6e-06;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 17 GCHGSEPCNIHRGKPF 32  
||||||| |||||||  
Db 20 GCHGSEPCIIHRGKPF 35

RESULT 14  
US-08-462-831-12  
; Sequence 12, Application US/08462831

```
; Patent No. 5552142
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
; TITLE OF INVENTION: DERMATOPHAGOIDES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,831
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/945,288
; FILING DATE: 10 SEPTEMBER 1992
; APPLICATION NUMBER: US 580,655
; FILING DATE: 11 SEPTEMBER 1990
; APPLICATION NUMBER: US 458,642
; FILING DATE: 13 FEBRUARY 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: MANDRAGOURAS, AMY E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 129 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 47
; OTHER INFORMATION: /label=Xaa is Thr or Ser
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 114
; OTHER INFORMATION: /label=Xaa is Asp or Asn
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 127
; OTHER INFORMATION: /label=Xaa is Ile or Leu
; US-08-462-831-12
;
; Query Match 50.8%; Score 92; DB 1; Length 129;
; Best Local Similarity 93.8%; Pred. No. 9.6e-06;
; Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; Qy 17 GCHGSEPCNIHRGKPF 32
; Db 20 GCHGSEPCNIHRGKPF 35
;
; RESULT 15
; US-08-461-809-12
; Sequence 12, Application US/08461809
; Patent No. 5770202
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
```

```
; TITLE OF INVENTION: DERMATOPHAGOIDES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,809
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/945,288
; FILING DATE: 10 SEPTEMBER 1992
; APPLICATION NUMBER: US 580,655
; FILING DATE: 11 SEPTEMBER 1990
; APPLICATION NUMBER: US 458,642
; FILING DATE: 13 FEBRUARY 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: MANDRAGOURAS, AMY E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 129 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 47
; OTHER INFORMATION: /label=Xaa is Thr or Ser
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 114
; OTHER INFORMATION: /label=Xaa is Asp or Asn
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 127
; OTHER INFORMATION: /label=Xaa is Ile or Leu
; US-08-461-809-12
;
; Query Match 50.8%; Score 92; DB 1; Length 129;
; Best Local Similarity 93.8%; Pred. No. 9.6e-06;
; Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; Qy 17 GCHGSEPCNIHRGKPF 32
; Db 20 GCHGSEPCNIHRGKPF 35
;
; Search completed: April 14, 2001, 10:16:56
; Job time: 441 sec
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2001, 10:13:16 ; Search time 77.31 Seconds  
(without alignments)  
48.514 Million cell updates/sec

Title: US-09-362-731-5

Perfect score: 181

Sequence: 1 QIIRANSFKIGITELGGCHGSEPCNIHRGKPF 32

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL15:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_unclassified:\*

13: sp\_vertebrate:\*

14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	38.1	135	5 Q9TZZ2	Q9TZZ2 euroglyphus
2	69	38.1	145	5 Q96430	Q96430 euroglyphus
3	53.5	29.6	403	10 O80919	O80919 arabidopsis
4	52	28.7	289	5 Q07338	Q07338 pyemotes tr
5	52	28.7	1884	11 P70207	P70207 mus musculus
6	51	28.2	149	7 Q31495	Q31495 oncorhynch
7	51	28.2	355	8 Q9TE80	Q9TE80 echinococcu
8	50.5	27.9	60	7 Q31585	Q31585 salmo salar
9	50.5	27.9	71	7 Q9XRJ9	Q9XRJ9 salvelinus
10	50.5	27.9	244	7 Q31590	Q31590 salmo salar
11	50	27.6	1067	13 Q9YHU6	Q9YHU6 xenopus lae
12	50	27.6	1328	4 Q9UIW1	Q9UIW1 homo sapien
13	50	27.6	1963	4 Q75051	Q75051 homo sapien
14	50	27.6	3680	5 Q9VR08	Q9VR08 drosophila
15	49.5	27.3	67	7 Q31578	Q31578 salmo salar
16	49.5	27.3	253	14 Q9YMU0	Q9YMU0 lymantria d
17	49	27.1	170	13 Q73893	Q73893 gallus gall
18	49	27.1	510	10 Q9LE20	Q9LE20 arabidopsis
19	48.5	26.8	202	8 Q9TIP4	Q9TIP4 orobanche h

20	48.5	26.8	204	8	O47109	O47109 orobanche r
21	48	26.5	62	2	Q9X5H5	Q9X5H5 helicobacte
22	48	26.5	484	1	O58041	O58041 pyrococcus
23	48	26.5	484	1	Q9UY44	Q9UY44 pyrococcus
24	48	26.5	502	10	O82752	O82752 arabidopsis
25	48	26.5	532	10	Q9SLV0	Q9SLV0 arabidopsis
26	48	26.5	817	5	Q21854	Q21854 caenorhabdi
27	47.5	26.2	173	10	O48704	O48704 arabidopsis
28	47.5	26.2	533	4	Q16518	Q16518 homo sapien
29	47.5	26.2	533	6	Q28175	Q28175 bos taurus
30	47.5	26.2	533	6	O97623	O97623 canis famil
31	47.5	26.2	533	6	O9XT71	O9XT71 cercopithe
32	47.5	26.2	533	6	Q05661	Q05661 bos taurus
33	47.5	26.2	533	6	Q9TVB8	Q9TVB8 canis famil
34	47.5	26.2	533	11	O70276	O70276 rattus norv
35	47.5	26.2	679	10	O9LPY5	O9LPY5 arabidopsis
36	47.5	26.2	1091	5	Q9W1Z6	Q9W1Z6 drosophila
37	47	26.0	38	2	O25251	O25251 helicobacte
38	47	26.0	147	4	Q9NWL9	Q9NWL9 homo sapien
39	47	26.0	158	2	Q9PRD7	Q9PRD7 ureaplasma
40	47	26.0	303	4	Q9NXX7	Q9NXX7 homo sapien
41	47	26.0	406	4	O9UNI9	O9UNI9 homo sapien
42	47	26.0	429	5	Q9V3S7	Q9V3S7 drosophila
43	47	26.0	630	14	Q65968	Q65968 cucurbit ap
44	46.5	25.7	67	7	Q31577	Q31577 salmo salar
45	46.5	25.7	67	7	Q31581	Q31581 salmo salar

#### ALIGNMENTS

RESULT 1

Q9TZZ2 PRELIMINARY; PRT; 135 AA.

AC Q9TZZ2: TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 15, Last annotation update)  
DE GROUP 2 ALLERGEN EUR M 2 0102 (FRAGMENT).  
GN EUR M 2 0102.

OS Euroglyphus maynei (House-dust mite).

OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;

OC Acariformes; Sarcophormes; Astigmata; Analgoidea; Pyroglyphidae;

OC Euroglyphus.

OX NCBI\_taxid=6958;

RN [1]

RP SEQUENCE FROM N.A.

RA Smith W., Hart B.J., Thomas W.R.;

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF047614; AAC82350.1; -

DR HSSP; P49278; 1A9V.

DR INTERPRO; IPR003172; -

DR PFAM; PF02221; EI\_DerP2\_DerF2; 1.

FT NON\_TER 1

SQ SEQUENCE 135 AA; 14555 MW; E15E1FA480B8DF16 CRC64;

Query Match 38.1%; Score 69; DB 5; Length 135;

Best Local Similarity 75.0%; Pred. No. 0.0088;

Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 17 GCHGSEPCNIHRGKPF 32

DB 26 GCKGSEPCVIHRGTAF 41

RESULT 2

Q96430 PRELIMINARY; PRT; 145 AA.

AC Q96430;

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

```

DE GROUP 2 ALLERGEN EUR M 2 0101.
GN EUR M 2 0101.
OS Euroglyphus maynei (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcophormes; Astigmata; Analgoidea; Pyroglyphidae;
OC Euroglyphus.
OX NCBI_TaxID=6958;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith W., Hart B.J., Thomas W.R.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF047613; AAC82349.1; -
DR HSSP; P49278; IASV.
DR INTERPRO; IPR003172; -
DR PFAM; PF02221; ElDerP2_DerF2; 1.
SQ SEQUENCE 145 AA; 15747 MW; 6655B16C8503A565 CRC64;

Query Match 38.1%; Score 69; DB 5; Length 145;
Best Local Similarity 75.0%; Pred. No. 0.0094;
Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 17 GCHGSEPCNIHRGKPF 32
Db 36 GCKGSEPCVIHRGTAF 51

RESULT 3
O80919 ID O80919 PRELIMINARY; PRT; 403 AA.
AC O80919;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DE T19C21.19 PROTEIN.
GN T19C21.19.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
RA Sykes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,
RA Somerville C.R., Venter J.C.;
RT "Arabidopsis thaliana chromosome II BAC T19C21 genomic sequence.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004683; AAC38772.1; -
SQ SEQUENCE 403 AA; 46816 MW; 564DB79AF14507E9 CRC64;

Query Match 29.6%; Score 53.5; DB 10; Length 403;
Best Local Similarity 44.4%; Pred. No. 5.9;
Matches 12; Conservative 4; Mismatches 10; Indels 1; Gaps 1;

Qy 4 KANSKFIGITELGGCHGSEPCNIHRGK 30
Db 33 KENPOSHGVTDGRGDSGRE-CNLFECK 58

RESULT 4
Q07338 ID Q07338 PRELIMINARY; PRT; 289 AA.
AC Q07338;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE INSECT-SELECTIVE NEUROTOXIN TXP-1 HOMOLOG PROTEIN.
GN T0X21A.
OS Pyemotes tritici (Straw itch mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;

```

---

```

OC Acariformes; Trombidiformes; Prostigmata; Anystina; Eleutherengona;
OC Heterostigmata; Pyemotoidea; Pyemotidae; Pyemotes.
OX NCBI_TaxID=6950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93227269; PubMed-8470135;
RA Tomalski M.D., Hutchinson K., Todd J., Miller L.K.;
RT "Identification and characterization of tox21A: a mite cDNA encoding a
RT paralytic neurotoxin related to Txp-1.";
RL Toxicon 31:319-326(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Tomalski M.D., Hutchinson K., Todd J., Miller L.K.;
RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; S58065; AAB26160.1; -
DR EMBL; L09272; AAA29801.1; -
KW Toxin; Neurotoxin.
SQ SEQUENCE 289 AA; 32731 MW; 476534ED1B84CB5B CRC64;

Query Match 28.7%; Score 52; DB 5; Length 289;
Best Local Similarity 41.9%; Pred. No. 7;
Matches 13; Conservative 4; Mismatches 10; Indels 4; Gaps 2;

Qy 4 KANSKF--IGITELGGCHGSEPCNIHRGKPF 32
Db 187 KNSKYEIYGVYLGRCGTG--NFSNGEVF 215

RESULT 5
P70207 ID P70207 PRELIMINARY; PRT; 1884 AA.
AC P70207;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE PLEXIN 2.
GN PLXN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE-96400270; PubMed-8806646;
RA Kameyama T., Murakami Y., Suto F., Kawakami A., Takagi S., Hirata T.,
RA Fujisawa H.;
RT "Identification of plexin family molecules in mice.";
RL Biochem. Biophys. Res. Commun. 226:396-402(1996).
DR EMBL; D86949; BAA13189.1; -
DR MGD; MGI:107684; Plxn2.
DR INTERPRO; IPR001627; -
DR INTERPRO; IPR002165; -
DR INTERPRO; IPR002909; -
DR PFAM; PF01403; Sema; 1.
DR PFAM; PF01437; Plexin_repeat; 3.
DR PFAM; PF01833; TIG; 4.
SQ SEQUENCE 1884 AA; 210181 MW; CC9BA9734BDCA703 CRC64;

Query Match 28.7%; Score 52; DB 11; Length 1884;
Best Local Similarity 37.9%; Pred. No. 51;
Matches 11; Conservative 4; Mismatches 10; Indels 4; Gaps 1;

Qy 3 IKANSKFIGITELGGCHGSEPCNIHRGK 31
Db 805 LKADHKF----ECGWCSEGRCTLHQHCP 829

RESULT 6
Q31495 ID Q31495 PRELIMINARY; PRT; 149 AA.

```



```

AC Q31495;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE MHC CLASS II BETA (FRAGMENT).
OS Oncorhynchus masou (Cherry salmon) (Masu salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8020;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BLOOD;
RX MEDLINE=96186525; PubMed=8606054;
RA Miller K.M., Withler R.E.;
RT "Sequence analysis of a polymorphic Mhc class II gene in Pacific
   salmon.";
RL Immunogenetics 43:337-351(1996).
DR EMBL: U34700; AAB01698.1; -.
DR INTERPRO: IPR000353; -.
DR INTERPRO: IPR003006; -.
DR PFAM: PF00047; Ig; 1.
DR PFAM: PF00969; MHC_II_beta; 1.
KW MHC.
FT NON_TER 1 1
FT NON_TER 149 149
SQ SEQUENCE 149 AA; 17003 MW; 5B56B3C2A74666EE CRC64;

Query Match 28.2%; Score 51; DB 7; Length 149;
Best Local Similarity 33.3%; Pred. No. 4.9;
Matches 14; Conservative 5; Mismatches 9; Indels 14; Gaps 2;

QY 1 QYIKANS---KFIGITELG-----GCHGSEPCNIHR 28
   :||: || :||: ||||| | | | | | | | |
Db 14 EYIRFNSTVGKFGVTELGKMNANWNSDAGILGQEQAELE 55

RESULT 7
Q9TE80 PRELIMINARY; PRT; 355 AA.
AC Q9TE80;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE CYTOCHROME B.
GN CYTB.
OS Echinococcus multilocularis.
OG Mitochondrion.
OC Eukaryota; Metazoa; Platyhelminthes; Neodermata; Cestoda; Eucestoda;
OC Cyclophyllidae; Taeniidae; Echinococcus.
OX NCBI_TaxID=6211;
RN [1]
RP SEQUENCE FROM N.A.
RA Fukunaga M.;
RT "Echinococcus multilocularis mitochondrial DNA sequence.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB018440; BA84928.2; -.
DR INTERPRO: IPR00179; -.
DR PFAM: PF00032; cytochrome_b_c; 1.
DR PFAM: PF00033; cytochrome_b_n; 1.
DR PROSITE: PS00193; CYTOCHROME_B_QQ; UNKNOWN_1.
KW Mitochondrion.
SQ SEQUENCE 355 AA; 41268 MW; C51D1A0EC7C6E948 CRC64;

Query Match 28.2%; Score 51; DB 8; Length 355;
Best Local Similarity 52.9%; Pred. No. 12;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 7 SKFIGITELGGCHGSEP 23
   | | :| ||||| |
Db 317 SLFLSLTYLGGCHPEYP 333

```

```

RESULT 8
Q31585 PRELIMINARY; PRT; 60 AA.
AC Q31585;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE (DB02) MHC CLASS II BETA 1 (FRAGMENT).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RA Grimholt U., Olsaker I., Vries Lindstrom C., Lie O.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL: L24953; AAA49597.1; -.
DR INTERPRO: IPR000353; -.
DR PFAM: PF00969; MHC_II_beta; 1.
KW MHC.
FT NON_TER 1 1
FT NON_TER 60 60
SQ SEQUENCE 60 AA; 6776 MW; 5D4F3449060940E2 CRC64;

Query Match 27.9%; Score 50.5; DB 7; Length 60;
Best Local Similarity 57.9%; Pred. No. 2.2;
Matches 11; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITELG 16
   :||: || ||: || ||||
Db 16 EYIRFNSTVGKFGVTELG 34

RESULT 9
Q9XRJ9 PRELIMINARY; PRT; 71 AA.
AC Q9XRJ9;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE MHC CLASS II BETA 1 (FRAGMENT).
GN SANA.
OS Salvelinus namaycush (lake trout).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salvelinus.
OX NCBI_TaxID=8040;
RN [1]
RP SEQUENCE FROM N.A.
RA Dorschner M.O., Duris T., Phillips R.B.;
RT "Diversity of a Lake Trout Mhc class II Gene.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF130026; AAD20889.1; -.
DR INTERPRO: IPR000353; -.
DR PFAM: PF00969; MHC_II_beta; 1.
FT NON_TER 1 1
FT NON_TER 71 71
SQ SEQUENCE 71 AA; 8002 MW; E3095286582A9F2D CRC64;

Query Match 27.9%; Score 50.5; DB 7; Length 71;
Best Local Similarity 57.9%; Pred. No. 2.6;
Matches 11; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITELG 16
   :||: || ||: || ||||
Db 14 EYIRFNSTVGKFGVTELG 32

```

```

RESULT 10
Q31590
ID Q31590 PRELIMINARY; PRT; 244 AA.
AC Q31590;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE MHC CLASS II.
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEUKOCYTES;
RA Hordvik I., Grimholt U., Fosse V.M., Lie Y., Endresen C.;
RL Immunogenetics 0-0-0(0).
DR EMBL; X70166; CAA49725.1; -
DR INTERPRO; IPR000353; -
DR PFAM; PF000047; Ig; 1.
DR PFAM; PF00969; MHC_II_beta; 1.
DR PRODOM; PD000328; -; 1.
KW MHC.
SQ SEQUENCE 244 AA; 27449 MW; 496CB9EA9D73765C CRC64;

Query Match 27.9%; Score 50.5; DB 7; Length 244;
Best Local Similarity 57.9%; Pred. No. 9.8;
Matches 11; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITELG 16
   :||: || ||: ||||
Db 51 EYIRENSTGKFGVGTGLG 69

RESULT 11
Q9YHU6
ID Q9YHU6 PRELIMINARY; PRT; 1067 AA.
AC Q9YHU6;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PHOSPHOLIPASE C-GAMMA-1B (FRAGMENT).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Kung H.-F., Kim J., Huang Y.K., Lin M.C.;
RT "Molecular cloning of two different forms of xenopus phospholipase C-
gamma-1.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF090112; AAD03595.1; -
DR HSSP; P08487; 2PLD
DR INTERPRO; IPR000008; -
DR INTERPRO; IPR000909; -
DR INTERPRO; IPR000980; -
DR INTERPRO; IPR001192; -
DR INTERPRO; IPR001452; -
DR INTERPRO; IPR001711; -
DR INTERPRO; IPR001849; -
DR INTERPRO; IPR002048; -
DR PFAM; PF00017; SH2; 2.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00168; C2; 1.
DR PFAM; PF00169; PH; 2.
DR PFAM; PF00387; PI-PLC-Y; 1.
DR PFAM; PF00388; PI-PLC-X; 1.
DR PRINTS; PR00390; PHPLIPASEC.

```

```

DR PRINTS; PR00401; SH2DOMAIN.
DR PRINTS; PR00452; SH3DOMAIN.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00001; SH2; 2.
DR PROSITE; PS00002; SH3; 1.
DR PROSITE; PS00003; PH_DOMAIN; 1.
DR PROSITE; PS00004; C2_DOMAIN_2; 1.
DR PROSITE; PS00007; PIPLC_X_DOMAIN; 1.
DR PROSITE; PS00008; PIPLC_Y_DOMAIN; 1.
FT NON_TER 1
FT NON_TER 1067
SQ SEQUENCE 1067 AA; 124019 MW; 533F6876ECDFE5DB CRC64;

Query Match 27.6%; Score 50; DB 13; Length 1067;
Best Local Similarity 31.8%; Pred. No. 56;
Matches 7; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 11 GITELGGCHGSEPCNIHRGKPF 32
   | : : | : : | : |
Db 745 GVLDVPSCHIAPRODVHNGRPF 766

RESULT 12
Q9UIW1
ID Q9UIW1 PRELIMINARY; PRT; 1328 AA.
AC Q9UIW1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE OCT/PLEXIN-A2 PROTEIN (FRAGMENT).
GN OCT/PLEXIN-A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKELETAL MUSCLE;
RX MEDLINE=96149362; PubMed=8570614;
RA Maestrini E., Tamagnone L., Longati P., Cremona O., Gullisano M.,
RA Bione S., Tamanini F., Neel B.G., Toniolo D., Comoglio P.M.;
RT "A family of transmembrane proteins with homology to the MET-
RT hepatocyte growth factor receptor.";
RN Proc. Natl. Acad. Sci. U.S.A. 93:674-678(1996).
RL [2]
RP SEQUENCE FROM N.A.
RC TISSUE=SKELETAL MUSCLE;
RA Tamagnone L., Artigiani S., Chen H., He Z., Ming G.L., Song H.L.,
RA Chedotal A., Winberg M.L., Goodman C.S., Poo M.M., Tessier-Lavigne M.,
RA Comoglio P.M.;
RT "Plexins are a large family of receptors for transmembrane, secreted
RT and GPI-anchored semaphorins in vertebrates.";
RL Cell 99:71-80(1999).
DR EMBL; X87831; CAB57275.1; -
DR INTERPRO; IPR001627; -
DR INTERPRO; IPR002165; -
DR INTERPRO; IPR002909; -
DR PFAM; PF01403; Sema; 1.
DR PFAM; PF01437; Plexin_repeat; 3.
DR PFAM; PF01833; TIG; 4.
FT NON_TER 1
FT NON_TER 1328
SQ SEQUENCE 1328 AA; 147724 MW; 434CF0F0FA42D8D9 CRC64;

Query Match 27.6%; Score 50; DB 4; Length 1328;
Best Local Similarity 38.5%; Pred. No. 70;
Matches 10; Conservative 4; Mismatches 8; Indels 4; Gaps 1;

QY 3 IKANSKFIGITELGGCHGSEPCNIHR 28
   :||: || | | | | :|:
Db 369 LKADRKPF---ECGWCSEGRRCCTLHQ 390

```

```

RESULT 13
O75051 PRELIMINARY; PRT; 1963 AA.
AC O75051;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DE K1AA0463 PROTEIN (FRAGMENT).
GN K1AA0463.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98116662; PubMed=9455484;
RA Seki N., Ohira M., Nagase T., Ishikawa K., Miyajima N., Nakajima D.,
RA Nomura N., Ohara O.;
RT "Characterization of cDNA clones in size-fractionated cDNA libraries
RT from human brain";
RL DNA Res. 4:345-349(1997).
DR EMBL; AB007932; BAA32308.1; -
DR INTERPRO; IPR001627; -
DR INTERPRO; IPR002165; -
DR INTERPRO; IPR002909; -
DR PFAM; PF01403; Sema; 1.
DR PFAM; PF01437; Plexin_repeat; 3.
DR PFAM; PF01833; TIG; 4.
FT NON_TER 1
SQ SEQUENCE 1963 AA; 218301 MW; 1C0AB9BA435F7804 CRC64;

Query Match 27.6%; Score 50; DB 4; Length 1963;
Best Local Similarity 38.5%; Pred. No. 1.1e+02;
Matches 10; Conservative 4; Mismatches 8; Indels 4; Gaps 1;

QY 3 IKANSKFIGITELGGCHGSEPCNIHR 28
Db 884 LXADRF-----ECGSGERRRCLHQ 905

RESULT 14
Q9VR08 PRELIMINARY; PRT; 3680 AA.
AC Q9VR08;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CG15637 PROTEIN.
GN CG15637.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

```

```

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003576; AAF51000.1; -
DR HSSP; P35555; IEMN.
DR FLYBASE; FBgn0031612; CG15637.
DR INTERPRO; IPR000152; -
DR INTERPRO; IPR000561; -
DR INTERPRO; IPR001881; -
DR PFAM; PF00008; EGF; 28.
DR PROSITE; PS00010; ASX-HYDROXYL; 18.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 18.
DR PROSITE; PS01187; EGF_CA; 17.
SQ SEQUENCE 3680 AA; 388637 MW; E5E972E1A3479EFF CRC64;

```

Query Match 27.6%; Score 50; DB 5; Length 3680;  
 Best Local Similarity 42.1%; Pred. No. 2.1e+02;  
 Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

```

QY 13 TELGGCHGSEPCNIHRGKP 31
:| | | :| | |
Db 745 SENGQCQDVDCSVQGRKP 763

```

```

RESULT 15
Q31578 PRELIMINARY; PRT; 67 AA.
AC Q31578;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE (DB03) MHC CLASS II BETA 1 (FRAGMENT).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RA Grimholt U., Olsaker I., Vries Lindstrom C., Lie O.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L24929; AAA49590.1; -
DR INTERPRO; IPR000353; -
DR PFAM; PF00969; MHC_II_beta; 1.
KW MHC.
FT NON_TER 1
FT NON_TER 67

```

SQ SEQUENCE 67 AA; 7449 MW; 42771AEDBABA6626 CRC64;

Query Match 27.3%; Score 49.5; DB 7; Length 67;  
Best Local Similarity 52.6%; Pred. No. 3.5;  
Matches 10; Conservative 4; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITELG 16

Db ::: || ||: ||||

16 EYVRFNSTVGKFGVGYTELG 34

Search completed: April 14, 2001, 10:13:19  
Job time: 224 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2001, 10:16:55 ; Search time 41.8 Seconds  
(without alignments)  
18.384 Million cell updates/sec

Title: US-09-362-731-4

Perfect score: 203

Sequence: 1 PKYVQNFLKATGKPKYVQNFLKATGKKGVIIGIK 40

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*

2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*

3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*

4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*

5: /cgn2\_6/ptodata/2/iaa/PCTUS.COMB.pep.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	81.5	40.1	347	1	US-08-329-781-58
2	81.5	40.1	347	1	US-08-630-918-58
3	80.5	39.7	566	4	US-09-232-468A-22
4	77	37.9	25	2	US-08-484-905-24
5	77	37.9	25	3	US-08-481-985B-24
6	77	37.9	25	4	US-08-370-476-24
7	75.5	37.2	570	2	US-08-453-848-7
8	75.5	37.2	571	2	US-08-453-848-15
9	75.5	37.2	571	2	US-08-453-848-21
10	72	35.5	14	3	US-09-125-517A-61
11	72	35.5	14	4	US-09-319-870A-9
12	72	35.5	15	3	US-09-125-517A-60
13	72	35.5	16	4	US-09-284-625-24
14	72	35.5	16	4	US-09-284-625-26
15	66	32.5	13	1	US-07-831-895C-5
16	66	32.5	13	1	US-08-305-871A-1
17	66	32.5	13	1	US-08-465-167A-22
18	66	32.5	13	2	US-08-480-190-24
19	66	32.5	13	2	US-08-596-387B-8
20	66	32.5	13	2	US-08-488-379-24
21	66	32.5	13	2	US-08-968-676-164
22	66	32.5	13	5	PCT-US92-07218-18
23	66	32.5	13	5	PCT-US93-07545-24
24	66	32.5	13	5	PCT-US94-10257A-20
25	66	32.5	13	5	PCT-US95-02121-99
26	66	32.5	13	5	PCT-US95-04121-53
27	66	32.5	13	5	PCT-US95-09816A-8

28	65	32.0	14	4	US-09-319-870A-8	Sequence 8, Appli
29	62	30.5	13	1	US-08-619-645-9	Sequence 9, Appli
30	62	30.5	13	2	US-08-634-493-9	Sequence 9, Appli
31	60	29.6	13	5	PCT-US95-04121-57	Sequence 57, Appli
32	60	29.6	13	5	PCT-US95-04121-58	Sequence 58, Appli
33	60	29.6	13	5	PCT-US95-04121-59	Sequence 59, Appli
34	55.5	27.3	566	4	US-09-232-468A-14	Sequence 14, Appli
35	52	25.6	348	1	US-08-229-781-50	Sequence 50, Appli
36	52	25.6	348	1	US-08-630-918-58	Sequence 50, Appli
37	49.5	24.4	1148	2	US-08-313-185-58	Sequence 58, Appli
38	49.5	24.4	1148	4	US-09-082-614A-58	Sequence 58, Appli
39	48	23.6	1475	3	US-09-007-999-2	Sequence 2, Appli
40	47	23.2	980	1	US-08-220-151-5	Sequence 5, Appli
41	47	23.2	980	1	US-08-413-118-5	Sequence 5, Appli
42	47	23.2	980	3	US-08-473-446-5	Sequence 5, Appli
43	47	23.2	2482	1	US-08-328-254-6	Sequence 6, Appli
44	47	23.2	3248	1	US-08-353-700-1	Sequence 1, Appli
45	47	23.2	3248	5	PCT-US95-16216-1	Sequence 1, Appli

#### ALIGNMENTS

RESULT 1  
US-08-229-781-58  
; Sequence 58, Application US/08229781  
; Patent No. 5589174  
; GENERAL INFORMATION:  
; APPLICANT: Yoshinobu OKUNO et al.  
; TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODY  
; NUMBER OF SEQUENCES: 58  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack  
; STREET: 805 Fifteenth Street, N.W., #700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/229,781  
; FILING DATE: April 19, 1994  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/054,016  
; FILING DATE: April 29, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren M. Cheek, Jr.  
; REGISTRATION NUMBER: 33,367  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-8850  
; TELEFAX:  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 347 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL:  
; ANTI-SENSE:  
; FRAGMENT TYPE:  
; ORIGINAL SOURCE:  
; ORGANISM:  
; STRAIN:  
; INDIVIDUAL ISOLATE:  
; DEVELOPMENTAL STAGE:



```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,905
; FILING DATE: 07-JUNE-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/801,818
; FILING DATE: 05-DEC-1991
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/792,473
; FILING DATE: 15-NOV-1991
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E. R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 03495.0106-03000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-484-905-24

Query Match 37.9%; Score 77; DB 2; Length 25;
Best Local Similarity 83.3%; Pred. No. 0.00018;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 18 PKYVKQNTLKLATGKKGKGV 35
| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 PKYVKQNTLKLATGMRNV 19

RESULT 5
US-08-481-985B-24
; Sequence 24, Application US/08481985B
; Patent No. 601146
; GENERAL INFORMATION:
; APPLICANT: Mottez, Estelle
; APPLICANT: Abastado, Jean-Pierre
; APPLICANT: Kourilsky, Philippe
; TITLE OF INVENTION: Altered Major Histocompatibility Complex
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,985B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/801,818
; FILING DATE: 05-DEC-1991
; CLASSIFICATION: 435
```

```
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/792,473
; FILING DATE: 15-NOV-1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0106-04000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-481-985B-24

Query Match 37.9%; Score 77; DB 3; Length 25;
Best Local Similarity 83.3%; Pred. No. 0.00018;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 18 PKYVKQNTLKLATGKKGKGV 35
| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 PKYVKQNTLKLATGMRNV 19

RESULT 6
US-08-370-476-24
; Sequence 24, Application US/08370476
; Patent No. 6153408
; GENERAL INFORMATION:
; APPLICANT: Mottez, Estelle
; APPLICANT: Abastado, Jean-Pierre
; APPLICANT: Kourilsky, Philippe
; APPLICANT: Lone, Yu-Chun
; APPLICANT: Ojcius, David
; APPLICANT: Casrouge, Armanda
; TITLE OF INVENTION: Altered Major Histocompatibility Complex
; NUMBER OF SEQUENCES: 127
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,476
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/117,575
; FILING DATE: 07-SEP-1993
; APPLICATION NUMBER: US 08/072,787
; FILING DATE: 06-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/801,818
; FILING DATE: 05-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/792,473
; FILING DATE: 15-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
```

```

; ; REFERENCE/DOCKET NUMBER: 05243.0001-010000
; ; TELECOMMUNICATION INFORMATION:
; ; TELEPHONE: 202-408-4000
; ; TELEFAX: 202-408-4400
; ; INFORMATION FOR SEQ ID NO: 24:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 25 amino acids
; ; TYPE: amino acid
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: peptide
US-08-370-476-24

```

Query Match 37.9%; Score 77; DB 4; Length 25;  
Best Local Similarity 83.3%; Pred. NO. 0.00018;  
Matches 15; Conservative 1; Mismatches 2; Indels

QY 18 PKYVKQNTLKLATGKKGV 35  
 |||  
 Db 2 PKYVKQNTLKLATGMRNV 19

```

RESULT 7
US-08-453-848-7
; Sequence 7, Application US/08453848
; Patent No. 5858368
; GENERAL INFORMATION:
; APPLICANT: Smith, Gale Eugene
; APPLICANT: Volvovitz, Franklin
; APPLICANT: Wilkinson, Bethanie Eident
; APPLICANT: Voznesensky, Andrei I.
; APPLICANT: Hackett, Craig Stanway
; TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
; TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea I. Pabet
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-3450

```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/453,848  
 FILING DATE: 30-MAY-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/120,607  
 FILING DATE: 13-SEPT-1993  
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: MGS101CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404)-873-8794  
TELEFAX: (404)-873-8795

INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 570 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOL% CULE TYPE: peptide  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 FRAGMENT TYPE: N-terminal

```

; ORIGINAL SOURCE:
; ORGANISM: Influenza virus
; INDIVIDUAL ISOLATE: A/Beijing/32/92 rHA
; FEATURE:
; NAME/KEY: AcNPV 61K protein signal sequence
; LOCATION: 1 to 18
; FEATURE:
; NAME/KEY: mature rHA
; LOCATION: 19 to 552
US-08-453-848-7

```

Query Match 37.2%; Score 75.5; DB 2; Length 570;  
Best Local Similarity 56.7%; Pred. No. 0.011;  
Matches 17; Conservative 4; Mismatches 8; Indels

QY 6 QNTLKLATGKKGPYVKQNTLKLATGKGV 35  
 || :| | :| | | | | | | :|  
 Db 315 QNVNRITYG-ACPRYVKQNTLKLATGMNV 343

```

RESULT      8
US-08-453-848-15
; Sequence 15, Application US/08453848
; Patent No. 5858368
; GENERAL INFORMATION:
; APPLICANT: Smith, Gale Eugene
; APPLICANT: Volvovitz, Franklin
; APPLICANT: Wilkinson, Bethanie Eident
; APPLICANT: Voznesensky, Andrei I.
; APPLICANT: Hackett, Craig Stanway
; TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
; HEMAGGLUTININ MULTIVALENT VACCINES
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-3450

```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,848  
FILING DATE: 30-MAY-1995

PRIOR APPLICATION DATA: 435  
 APPLICATION NUMBER: 08/120,607  
 FILING DATE: 13-SEPT-1993  
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: MGS101CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404)-873-8794  
TELEFAX: (404)-873-8795

INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 571 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:



;; ORGANISM: Influenza virus  
;; INDIVIDUAL ISOLATE: A/Shandong/9/93 rHA  
;; FEATURE:  
;; NAME/KEY: ACPV 61K protein signal sequence  
;; LOCATION: 1 to 18  
;; FEATURE:  
;; NAME/KEY: mature rHA  
;; LOCATION: 19 to 553  
US-08-453-848-15

Query Match 37.2%; Score 75.5; DB 2; Length 571;  
Best Local Similarity 56.7%; Pred. No. 0.011;  
Matches 17; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

QY 6 QNTLKLATGKGPYVKQNTLKLATGKGV 35  
|| : : | : ||||| : : |  
DB 316 QNVNRTYG-ACPRYVKQNTLKLATGMRNV 344

RESULT 9  
US-08-453-848-21  
; Sequence 21, Application US/08453848  
; Patent No. 5858368  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Gale Eugene  
; APPLICANT: Volnovitz, Franklin  
; APPLICANT: Wilkinson, Bethanie Eident  
; APPLICANT: Voznesensky, Andrei I.  
; APPLICANT: Hackett, Craig Stanway  
; TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA  
; TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; STREET: 1201 West Peachtree Street  
; CITY: Atlanta  
; STATE: GA  
; COUNTRY: USA  
; ZIP: 30309-3450

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,848  
FILING DATE: 30-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/120,607  
FILING DATE: 13-SEPT-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: MGS101C1P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404)-873-8794  
TELEFAX: (404)-873-8795  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 571 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Influenza virus

;; INDIVIDUAL ISOLATE: A/Johannesburg/33/94 rHA  
;; FEATURE:  
;; NAME/KEY: ACPV 61K protein signal sequence  
;; LOCATION: 1 to 18  
;; FEATURE:  
;; NAME/KEY: mature rHA  
;; LOCATION: 19 to 569  
US-08-453-848-21

Query Match 37.2%; Score 75.5; DB 2; Length 571;  
Best Local Similarity 56.7%; Pred. No. 0.011;  
Matches 17; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

QY 6 QNTLKLATGKGPYVKQNTLKLATGKGV 35  
|| : : | : ||||| : : |  
DB 316 QNVNRTYG-ACPRYVKQNTLKLATGMRNV 344

RESULT 10  
US-09-125-517A-61  
; Sequence 61, Application US/09125517A  
; Patent No. 6087336  
; GENERAL INFORMATION:  
; APPLICANT: Cotton, Ronald  
; APPLICANT: Edwards, Philip Neil  
; APPLICANT: Luke, Richard William Arthur  
; TITLE OF INVENTION: Peptide Derivatives  
; NUMBER OF SEQUENCES: 62  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.  
; STREET: 555 Thirteenth St., N.W., Suite 701 East  
; STREET: Tower  
; CITY: Washington  
; STATE: DC  
; COUNTRY: U.S.A.  
; ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/125,517A  
FILING DATE: 20-AUG-1998  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9603855.9  
FILING DATE: 23-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9620819.4  
FILING DATE: 05-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Ernst, Barbara G.  
REGISTRATION NUMBER: 30,377  
REFERENCE/DOCKET NUMBER: 1991-127  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-783-6040  
TELEFAX: 202-783-6031  
INFORMATION FOR SEQ ID NO: 61:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-125-517A-61

Query Match 35.5%; Score 72; DB 3; Length 14;  
Best Local Similarity 100.0%; Pred. No. 0.00046;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKYVKONTLKLATG 14  
|||||  
DB 1 PKYVKONTLKLATG 14

## RESULT 11

US-09-319-870A-9  
; Sequence 9, Application US/09319870A  
; Patent No. 6184207

## GENERAL INFORMATION:

; APPLICANT: Astrazeneca  
; APPLICANT: Luke, Richard W. A.  
; APPLICANT: Cotton, Richard  
; TITLE OF INVENTION: Inhibitors of Peptide Binding to MHC Class II Proteins  
; FILE REFERENCE: 1991-175  
; CURRENT APPLICATION NUMBER: US/09/319,870A  
; CURRENT FILING DATE: 1999-06-14  
; PRIOR APPLICATION NUMBER: PCT/GB97/03397  
; PRIOR FILING DATE: 1996-12-12  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 14  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Modified  
; OTHER INFORMATION: Sequence  
US-09-319-870A-9

Query Match 35.5%; Score 72; DB 4; Length 14;  
Best Local Similarity 100.0%; Pred. No. 0.00046;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKYVKONTLKLATG 14  
|||||  
DB 1 PKYVKONTLKLATG 14

## RESULT 12

US-09-125-517A-60  
; Sequence 60, Application US/09125517A  
; Patent No. 6087336

## GENERAL INFORMATION:

; APPLICANT: Cotton, Ronald  
; APPLICANT: Edwards, Philip Neil  
; APPLICANT: Luke, Richard William Arthur  
; TITLE OF INVENTION: Peptide Derivatives  
; NUMBER OF SEQUENCES: 62  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Rothwell, Figg, Ernst & Kurz, P.C.  
; STREET: 555 Thirteenth St., N.W., Suite 701 East  
; CITY: Washington  
; STATE: DC  
; COUNTRY: U.S.A.  
; ZIP: 20004

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/125,517A  
; FILING DATE: 20-AUG-1998  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9603855.9  
; FILING DATE: 23-FEB-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9620819.4  
; FILING DATE: 05-OCT-1996

; ATTORNEY/AGENT INFORMATION:  
; NAME: Ernst, Barbara G.  
; REGISTRATION NUMBER: 30,377  
; REFERENCE/DOCKET NUMBER: 1991-127  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-783-6040  
; TELEFAX: 202-783-6031  
; INFORMATION FOR SEQ ID NO: 60:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 1  
; OTHER INFORMATION: /product= "OTHER"  
; OTHER INFORMATION: /note= "biotinyl-2-aminohexanoyl"  
US-09-125-517A-60

Query Match 35.5%; Score 72; DB 3; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.0005;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKYVKONTLKLATG 14  
|||||  
DB 2 PKYVKONTLKLATG 15

## RESULT 13

US-09-284-625-24  
; Sequence 24, Application US/09284625  
; Patent No. 6207644

## GENERAL INFORMATION:

; APPLICANT: Luke, Richard William  
; APPLICANT: Cotton, Richard  
; TITLE OF INVENTION: Peptide Analogues Containing a 7-membered Lactam Ring  
; FILE REFERENCE: 1991-169  
; CURRENT APPLICATION NUMBER: US/09/284,625  
; CURRENT FILING DATE: 1999-04-16  
; PRIOR APPLICATION NUMBER: PCT/GB97/02837  
; PRIOR FILING DATE: 1997-10-14  
; PRIOR APPLICATION NUMBER: GB 96218367  
; PRIOR FILING DATE: 1996-10-19  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (1)  
; OTHER INFORMATION: Biotin-Ahx  
; NAME/KEY: MOD\_RES  
; LOCATION: (16)  
; OTHER INFORMATION: -OH  
US-09-284-625-24

Query Match 35.5%; Score 72; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.00053;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKYVKONTLKLATG 14  
|||||  
DB 2 PKYVKONTLKLATG 15

## RESULT 14

US-09-284-625-26

; Sequence 26, Application US/09284625  
; Patent No. 6207644  
; GENERAL INFORMATION:  
; APPLICANT: Luke, Richard William  
; TITLE OF INVENTION: Peptide Analogues Containing a 7-membered Lactam Ring  
; FILE REFERENCE: 1991-169  
; CURRENT APPLICATION NUMBER: US/09/284,625  
; PRIOR FILING DATE: 1999-04-16  
; PRIOR APPLICATION NUMBER: PCT/GB97/02837  
; PRIOR FILING DATE: 1997-10-14  
; PRIOR APPLICATION NUMBER: GB 96218367  
; PRIOR FILING DATE: 1996-10-19  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 26  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (1)  
; OTHER INFORMATION: -H  
; NAME/KEY: MOD\_RES  
; LOCATION: (16)  
; OTHER INFORMATION: -OH  
; US-09-284-625-26

Query Match 35.5%; Score 72; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.00053;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKYVKONTLKLATG 14  
| | | | | | | | | | | | | | | |  
Db 2 PKYVKONTLKLATG 15

RESULT 15  
US-07-831-895C-5  
; Sequence 5, Application US/07831895C  
; Patent No. 5583031  
; GENERAL INFORMATION:  
; APPLICANT: Stern, Lawrence J.  
; TITLE OF INVENTION: EMPTY MAJOR HISTOCOMPATIBILITY CLASS II  
; TITLE OF INVENTION: HETERODIMERS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/831,895C  
; FILING DATE: 06-FEB-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00246/139001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 13 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-831-895C-5  
  
Query Match 32.5%; Score 66; DB 1; Length 13;  
Best Local Similarity 100.0%; Pred. No. 0.0028;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PKYVKONTLKLAT 13  
| | | | | | | | | | | | | |  
Db 1 PKYVKONTLKLAT 13  
  
Search completed: April 14, 2001, 10:16:56  
Job time: 441 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2001, 10:13:14 ; Search time 77.31 Seconds  
(without alignments)  
60.643 Million cell updates/sec

Title: US-09-362-731-4

Perfect score: 203

Sequence: 1 PKYKQNTLKLATGKPKYKQNTLKLATGKKGVIIGIK 40

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_15:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_unclassified:\*

13: sp\_vertebrate:\*

14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85.5	42.1	344	14 Q86899	Q86899 equine infl
2	85.5	42.1	359	14 Q9YJW0	Q9YJW0 influenza a
3	85.5	42.1	565	14 Q82559	Q82559 influenza a
4	84.5	41.6	344	14 Q82847	Q82847 influenza a
5	82.5	40.6	332	14 Q56964	Q56964 influenza a
6	82.5	40.6	335	14 Q91205	Q91205 influenza a
7	82.5	40.6	342	14 Q56963	Q56963 influenza a
8	82.5	40.6	343	14 Q9YJW7	Q9YJW7 influenza a
9	82.5	40.6	344	14 Q82843	Q82843 influenza a
10	82.5	40.6	344	14 Q82844	Q82844 influenza a
11	82.5	40.6	344	14 Q82846	Q82846 influenza a
12	82.5	40.6	344	14 Q82848	Q82848 influenza a
13	82.5	40.6	344	14 Q906P1	Q906P1 equine infl
14	82.5	40.6	344	14 Q906P0	Q906P0 equine infl
15	82.5	40.6	344	14 Q906N7	Q906N7 equine infl
16	82.5	40.6	344	14 Q906N6	Q906N6 equine infl
17	82.5	40.6	344	14 Q906N5	Q906N5 equine infl
18	82.5	40.6	346	14 Q56857	Q56857 equine infl
19	82.5	40.6	346	14 Q56858	Q56858 equine infl

20	82.5	40.6	353	14 Q12294	Q12294 equine infl
21	82.5	40.6	357	14 Q56856	Q56856 equine infl
22	82.5	40.6	360	14 Q9YJW8	Q9YJW8 influenza a
23	82.5	40.6	360	14 Q9YJV3	Q9YJV3 influenza a
24	82.5	40.6	362	14 Q9YJV9	Q9YJV9 influenza a
25	82.5	40.6	362	14 Q9YJV6	Q9YJV6 influenza a
26	82.5	40.6	363	14 Q9YJV4	Q9YJV4 influenza a
27	82.5	40.6	365	14 Q56961	Q56961 influenza a
28	82.5	40.6	365	14 Q56962	Q56962 influenza a
29	82.5	40.6	365	14 Q86639	Q86639 equine infl
30	82.5	40.6	365	14 Q82792	Q82792 influenza a
31	82.5	40.6	365	14 Q82793	Q82793 influenza a
32	82.5	40.6	365	14 Q67103	Q67103 equine infl
33	82.5	40.6	365	14 Q67104	Q67104 equine infl
34	82.5	40.6	365	14 Q67105	Q67105 equine infl
35	82.5	40.6	365	14 Q67106	Q67106 equine infl
36	82.5	40.6	365	14 Q67107	Q67107 equine infl
37	82.5	40.6	365	14 Q9WA90	Q9WA90 influenza a
38	81.5	40.1	328	14 Q90384	Q90384 influenza a
39	81.5	40.1	329	14 Q67024	Q67024 influenza a
40	81.5	40.1	329	14 Q67025	Q67025 influenza a
41	81.5	40.1	329	14 Q67342	Q67342 influenza a
42	81.5	40.1	329	14 Q67346	Q67346 influenza a
43	81.5	40.1	329	14 Q67351	Q67351 influenza a
44	81.5	40.1	329	14 Q67352	Q67352 influenza a
45	81.5	40.1	329	14 Q67353	Q67353 influenza a

#### ALIGNMENTS

RESULT 1

ID Q86899 PRELIMINARY; PRT; 344 AA.  
AC Q86899;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
DE HEMAGGLUTININ (FRAGMENT).  
GN HAL.  
OS Equine influenza virus.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group.  
OX NCBI\_TaxID=11312;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95159661; PubMed=7545975;  
RA Oxburgh L., Berg M., Klingeborn B., Emmoth E., Linne T.;  
RT "Evolution of H3N8 equine influenza virus from 1963 to 1991."  
RL Virus Res. 34:153-165(1994).  
DR EMBL; S77429; AAB33340.1; -.  
DR INTERPRO; IPR000149; -.  
DR INTERPRO; IPR001364; -.  
DR PRINTS; PR00330; HEMAGGLUTN1.  
ET NON\_TER 344 344  
SQ SEQUENCE 344 AA; 38284 MW; C885938341149DA6 CRC64;

Query Match 42.1%; Score 85.5; DB 14; Length 344;  
Best\_Local Similarity 63.3%; Pred. No. 0.0047; Indels 1; Gaps 1;  
Matches 19; Conservative 3; Mismatches 7;

Qy 6 QNTLKLATGKPKYKQNTLKLATGKKG 35

Db 310 QNVNKTG-RCPKYIKNTLKLATGMNV 338

RESULT 2

ID Q9YJW0 PRELIMINARY; PRT; 359 AA.  
AC Q9YJW0;  
DT 01-MAY-1999 (Tremblrel. 10, Created)

```

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE HEMAGGLUTININ PRECURSOR (FRAGMENT)
OS Influenza A virus (strain A/Mallard/Ohio/556/1987)
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
OX NCBI_TaxID=86067;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H3N8;
RA Lindstrom S.E., Endo A., Pecoraro M.R., Sugita S., Damiani A.,
RA Hiromoto Y., Kanata K., Kumamoto T., Nerome K.;
RT "Complete nucleotide sequence of the HA1 region of the hemagglutinin
RL gene of A/Equine/Newmarket/D64/79 (H3N8) influenza virus.";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D30677; BAA33938.1; -
DR INTERPRO; IPR000149; -
DR INTERPRO; IPR001364; -
DR PFAM; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
KW SIGNAL.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 >343 HEMAGGLUTININ 1 (HA1).
FT NON_TER 359 359
SO SEQUENCE 359 AA; 39900 MW; E12B17A7A02C75D4 CRC64;

Query Match 42.1%; Score 85.5; DB 14; Length 359;
Best Local Similarity 63.3%; Pred. No. 0.0049; Mismatches 7; Indels 1; Gaps 1;
Matches 19; Conservative 3;

Qy 6 QNTLKATGKKGPKYKQNTLKATGKKG 35
Db 310 QNVNKITYG-KCPKYKQNTLKATGMRNV 338

RESULT 3
Q82559 ID 082559 PRELIMINARY; PRT; 565 AA.
AC 082559;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE HEMAGGLUTININ.
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/EQUINE/KENTUCKY/1/81;
RA Olsen C.W., MacGregor M.W., Hinshaw V.S.;
RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U58195; AAB02560.1; -
DR INTERPRO; IPR000149; -
DR INTERPRO; IPR000386; -
DR INTERPRO; IPR001364; -
DR PFAM; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
DR PRINTS; PR00331; HEMAGGLUTN2.
DR PRODOM; PD000225; -; 1.
SO SEQUENCE 565 AA; 63633 MW; E8412D13945FA424 CRC64;

Query Match 42.1%; Score 85.5; DB 14; Length 565;
Best Local Similarity 63.3%; Pred. No. 0.0082;
Matches 19; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

Qy 6 QNTLKATGKKGPKYKQNTLKATGKKG 35
Db 310 QNVNKITYG-KCPKYKQNTLKATGMRNV 338

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE HEMAGGLUTININ PRECURSOR (FRAGMENT)
OS Influenza A virus (strain A/Mallard/Ohio/556/1987)
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
OX NCBI_TaxID=86067;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H3N8;
RA Lindstrom S.E., Endo A., Pecoraro M.R., Sugita S., Damiani A.,
RA Hiromoto Y., Kanata K., Kumamoto T., Nerome K.;
RT "Complete nucleotide sequence of the HA1 region of the hemagglutinin
RL gene of A/Equine/Newmarket/D64/79 (H3N8) influenza virus.";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D30677; BAA33938.1; -
DR INTERPRO; IPR000149; -
DR INTERPRO; IPR001364; -
DR PFAM; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
KW SIGNAL.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 >343 HEMAGGLUTININ 1 (HA1).
FT NON_TER 359 359
SO SEQUENCE 359 AA; 39900 MW; E12B17A7A02C75D4 CRC64;

Query Match 42.1%; Score 85.5; DB 14; Length 359;
Best Local Similarity 63.3%; Pred. No. 0.0049; Mismatches 7; Indels 1; Gaps 1;
Matches 19; Conservative 3;

Qy 6 QNTLKATGKKGPKYKQNTLKATGKKG 35
Db 310 QNVNKITYG-KCPKYKQNTLKATGMRNV 338

RESULT 3
Q82559 ID 082559 PRELIMINARY; PRT; 565 AA.
AC 082559;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE HEMAGGLUTININ.
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/EQUINE/KENTUCKY/1/81;
RA Olsen C.W., MacGregor M.W., Hinshaw V.S.;
RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U58195; AAB02560.1; -
DR INTERPRO; IPR000149; -
DR INTERPRO; IPR000386; -
DR INTERPRO; IPR001364; -
DR PFAM; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
DR PRINTS; PR00331; HEMAGGLUTN2.
DR PRODOM; PD000225; -; 1.
SO SEQUENCE 565 AA; 63633 MW; E8412D13945FA424 CRC64;

Query Match 42.1%; Score 85.5; DB 14; Length 565;
Best Local Similarity 63.3%; Pred. No. 0.0082;
Matches 19; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

Qy 6 QNTLKATGKKGPKYKQNTLKATGKKG 35
Db 310 QNVNKITYG-KCPKYKQNTLKATGMRNV 338

```

```

RESULT 4
Q82847 ID 082847 PRELIMINARY; PRT; 344 AA.
AC 082847;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE HAY SUBUNIT OF HAEMAGGLUTININ (FRAGMENT).
GN HAI.
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEWMARKET/93, CV. H3N8(EQUINE);
RA Daly J.M., Lai A.C., Binns M.M., Chambers T.M., Barrandeguy M.,
RA Mumford J.A.;
RT "Antigenic and genetic evolution of equine H3N8 influenza A viruses.";
RL J. Gen. Virol. 77:661-671(1996).
DR EMBL; X85089; CAA59416.2; -
DR INTERPRO; IPR000149; -
DR INTERPRO; IPR001364; -
DR PFAM; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
FT NON_TER 344 344
SO SEQUENCE 344 AA; 38270 MW; 4E271E05DDB6A4910 CRC64;

Query Match 41.6%; Score 84.5; DB 14; Length 344;
Best Local Similarity 63.3%; Pred. No. 0.0063;
Matches 19; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

Qy 6 QNTLKATGKKGPKYKQNTLKATGKKG 35
Db 310 QNVNKITYG-KCPKYKQNTLKATGMRNV 338

RESULT 5
Q56964 ID 056964 PRELIMINARY; PRT; 332 AA.
AC 056964;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE HAEMAGGLUTININ (FRAGMENT).
GN HA.
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/EQUINE/BERLIN/1/91(H3N8);
RA Ilobi C.P., Nicolson C., Taylor J., Mumford J.A., Wood J.M.,
RA Robertson J.S.;
RA Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ223196; CAA11171.1; -
DR INTERPRO; IPR000149; -
DR INTERPRO; IPR001364; -
DR PFAM; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00330; HEMAGGLUTN1.
FT NON_TER 1 1
FT CHAIN 1 >328 HAI HAEMAGGLUTININ.
FT CHAIN 330 >332 HA2 HAEMAGGLUTININ.
FT NON_TER 332 332
SO SEQUENCE 332 AA; 36858 MW; 82C9E124C47933E7 CRC64;

Query Match 40.6%; Score 82.5; DB 14; Length 332;

```

```
Best Local Similarity 60.0%; Pred. No. 0.011;
Matches 18; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

QY 6 QNTLKLATGKGGPKYVKQNTLKLATGKGV 35
   || | | | | | | | | | | | | | | | |
Db 295 QNVNKVTYG-KCPKYIRQNTLKLATGMRNV 323

RESULT 6
O91205 PRELIMINARY; PRT; 335 AA.
AC O91205;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE HAEMAGGLUTININ (FRAGMENT).
GN HA.
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/EQUINE/BERLIN/4/89(H3N8);
RX MEDLINE=98309066; PubMed=9645196;
RA Illobi C.P., Nicolson C., Taylor J., Mumford J.A., Wood J.M.,
RA Robertson J.S.;
RA "Direct sequencing of the HA gene of clinical equine H3N8 influenza
RT virus and comparison with laboratory derived viruses.";
RL Arch. Virol. 143:891-901(1998).
DR EMBL; AJ223195; CAAL1170.1; -.
DR INTERPRO; IPR000149; -.
DR INTERPRO; IPR001364; -.
DR PFAM; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00330; HEMAGGLUTN1.
DR NON_TER 1
FT CHAIN 1 >328 HA1 HAEMAGGLUTININ.
FT CHAIN 330 >335 HA2 HAEMAGGLUTININ.
FT NON_TER 335
SQ SEQUENCE 335 AA; 37055 MW; 2D54B5A8E14782D8 CRC64;

Query Match 40.6%; Score 82.5; DB 14; Length 335;
Best Local Similarity 60.0%; Pred. No. 0.011;
Matches 18; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

QY 6 QNTLKLATGKGGPKYVKQNTLKLATGKGV 35
   || | | | | | | | | | | | | | | | |
Db 295 QNVNKVTYG-KCPKYIRQNTLKLATGMRNV 323

RESULT 7
O56963 PRELIMINARY; PRT; 342 AA.
ID O56963;
AC O56963;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE HAEMAGGLUTININ (FRAGMENT).
GN HA.
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/EQUINE/BERLIN/3/89(H3N8);
RA Illobi C.P., Nicolson C., Taylor J., Mumford J.A., Wood J.M.,
RA Robertson J.S.;
RA Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RL EMBL; AJ223194; CAAL1169.1; -.
DR INTERPRO; IPR000149; -.
DR INTERPRO; IPR001364; -.

Query Match 40.6%; Score 82.5; DB 14; Length 342;
Best Local Similarity 60.0%; Pred. No. 0.011;
Matches 18; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

QY 6 QNTLKLATGKGGPKYVKQNTLKLATGKGV 35
   || | | | | | | | | | | | | | | | |
Db 295 QNVNKVTYG-KCPKYIRQNTLKLATGMRNV 323

RESULT 8
O9YJV7 PRELIMINARY; PRT; 343 AA.
ID O9YJV7;
AC O9YJV7;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE HEMAGGLUTININ PRECURSOR (FRAGMENT).
OS Influenza A virus (strain A/Mallard/Ohio/556/1987).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
OX NCBI_TaxID=86067;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H3N8;
RA Lindstrom S.E., Endo A., Pecoraro M.R., Sugita S., Damiani A.,
RA Hiromoto Y., Kamata M., Kumanomido T., Nerome K.;
RT "Complete nucleotide sequence of the HA1 region of the hemagglutinin
RL gene of A/Equine/Alaska/1/91 (H3N8) influenza virus.";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; D30680; BAA33941.1; -.
DR INTERPRO; IPR000149; -.
DR INTERPRO; IPR001364; -.
DR PFAM; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
KW Signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 >343 HEMAGGLUTININ 1 (HA1).
FT NON_TER 343
SQ SEQUENCE 343 AA; 38108 MW; E77844BCEB899B0C CRC64;

Query Match 40.6%; Score 82.5; DB 14; Length 343;
Best Local Similarity 60.0%; Pred. No. 0.011;
Matches 18; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

QY 6 QNTLKLATGKGGPKYVKQNTLKLATGKGV 35
   || | | | | | | | | | | | | | | | |
Db 310 QNVNKVTYG-KCPKYIRQNTLKLATGMRNV 338

RESULT 9
O82843 PRELIMINARY; PRT; 344 AA.
ID O82843;
AC O82843;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE HAY SUBUNIT OF HAEMAGGLUTININ (FRAGMENT).
GN HAI.
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
OX NCBI_TaxID=11320;
```

```

RN RP SEQUENCE FROM N.A.
RC STRAIN=ARUNDEL /91, CV. H3N8(EQUINE);
RA Dally J.M., Lai A.C., Binns M.M., Chambers T.M., Barrandeguy M.,
RA Mumford J.A.;
RT "Antigenic and genetic evolution of equine H3N8 influenza A viruses.";
RL J. Gen. Virol. 77:661-671(1996).
DR EMBL; X85085; CAA59412.2; -.
DR INTERPRO; IPR000149; -.
DR INTERPRO; IPR001364; -.
DR PFAM; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
FT NON_TER 344
SQ SEQUENCE 344 AA; 38321 MW; B6BDCF71A7A3138E CRC64;

Query Match 40.6%; Score 82.5; DB 14; Length 344;
Best Local Similarity 60.0%; Pred. No. 0.011; 7; Indels 1; Gaps 1;
Matches 18; Conservative 4; Mismatches 7;

QY 6 QNTLKLATGKKGPKYVKQNTLKLATGKKG 35
Db 310 QNVNKVTYG-KCPKYIRQNTLKLATGMNV 338

RESULT 10
Q82844 PRELIMINARY; PRT; 344 AA.
AC Q82844;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HAY SUBUNIT OF HAEMAGGLUTININ (FRAGMENT).
GN HAI.
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ELLA /89, CV. H3N8(EQUINE);
RA Dally J.M., Lai A.C., Binns M.M., Chambers T.M., Barrandeguy M.,
RA Mumford J.A.;
RT "Antigenic and genetic evolution of equine H3N8 influenza A viruses.";
RL J. Gen. Virol. 77:661-671(1996).
DR EMBL; X85086; CAA59413.2; -.
DR INTERPRO; IPR000149; -.
DR INTERPRO; IPR001364; -.
DR PFAM; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
FT NON_TER 344
SQ SEQUENCE 344 AA; 38255 MW; 4C7C3192826633F9 CRC64;

Query Match 40.6%; Score 82.5; DB 14; Length 344;
Best Local Similarity 60.0%; Pred. No. 0.011; 7; Indels 1; Gaps 1;
Matches 18; Conservative 4; Mismatches 7;

QY 6 QNTLKLATGKKGPKYVKQNTLKLATGKKG 35
Db 310 QNVNKVTYG-KCPKYIRQNTLKLATGMNV 338

RESULT 11
Q82846 PRELIMINARY; PRT; 344 AA.
AC Q82846;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HAY SUBUNIT OF HAEMAGGLUTININ (FRAGMENT).

```

```

GN HAI.
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEWMARKET /93, CV. H3N8(EQUINE);
RA Dally J.M., Lai A.C., Binns M.M., Chambers T.M., Barrandeguy M.,
RA Mumford J.A.;
RT "Antigenic and genetic evolution of equine H3N8 influenza A viruses.";
RL J. Gen. Virol. 77:661-671(1996).
DR EMBL; X85088; CAA59415.2; -.
DR INTERPRO; IPR000149; -.
DR INTERPRO; IPR001364; -.
DR PFAM; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
FT NON_TER 344
SQ SEQUENCE 344 AA; 38295 MW; 3B167053FD8B53C1 CRC64;

Query Match 40.6%; Score 82.5; DB 14; Length 344;
Best Local Similarity 60.0%; Pred. No. 0.011; 7; Indels 1; Gaps 1;
Matches 18; Conservative 4; Mismatches 7;

QY 6 QNTLKLATGKKGPKYVKQNTLKLATGKKG 35
Db 310 QNVNKVTYG-KCPKYIRQNTLKLATGMNV 338

RESULT 12
Q82848 PRELIMINARY; PRT; 344 AA.
AC Q82848;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HAY SUBUNIT OF HAEMAGGLUTININ (FRAGMENT).
GN HAI.
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SUSSEX /89, CV. H3N8(EQUINE);
RA Dally J.M., Lai A.C., Binns M.M., Chambers T.M., Barrandeguy M.,
RA Mumford J.A.;
RT "Antigenic and genetic evolution of equine H3N8 influenza A viruses.";
RL J. Gen. Virol. 77:661-671(1996).
DR EMBL; X85090; CAA59417.2; -.
DR INTERPRO; IPR000149; -.
DR INTERPRO; IPR001364; -.
DR PFAM; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
FT NON_TER 344
SQ SEQUENCE 344 AA; 38281 MW; 65B76C4DBDBAFID5 CRC64;

Query Match 40.6%; Score 82.5; DB 14; Length 344;
Best Local Similarity 60.0%; Pred. No. 0.011; 7; Indels 1; Gaps 1;
Matches 18; Conservative 4; Mismatches 7;

QY 6 QNTLKLATGKKGPKYVKQNTLKLATGKKG 35
Db 310 QNVNKVTYG-KCPKYIRQNTLKLATGMNV 338

RESULT 13
Q906P1 PRELIMINARY; PRT; 344 AA.
ID Q906P1

```



```

AC Q906P1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HEMAGGLUTININ PRECURSOR (FRAGMENT).
GN HAI.
OS Equine influenza virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
OX NCBI_TaxID=11312;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A/EQ/FLORIDA/1/94;
RA Lai A.C.K.;
RT "Diverged evolution of equine-2 influenza (H3N8) virus in the Western Hemisphere.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF197242; AAF22346.1; -.
DR INTERPRO; IPR000149; -.
DR PFAM; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
KW Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 >344 HEMAGGLUTININ.
FT NON_TER 344 344
SQ SEQUENCE 344 AA; 38270 MW; 2A387D226D45880D CRC64;

Query Match 40.6%; Score 82.5; DB 14; Length 344;
Best Local Similarity 60.0%; Pred. No. 0.011;
Matches 18; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

QY 6 QNTLKATGKKGPKYVKQNTLKATGKGV 35
|| | : | | | : | | | | | | | : |
Db 310 QNVNKVTYG-KCPKYIRQNTLKATGMNV 338

RESULT 14
Q906P0 PRELIMINARY; PRT; 344 AA.
AC Q906P0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HEMAGGLUTININ PRECURSOR (FRAGMENT).
GN HAI.
OS Equine influenza virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
OX NCBI_TaxID=11312;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A/EQ/SASKATOON/1/90;
RA Lai A.C.K.;
RT "Diverged evolution of equine-2 influenza (H3N8) virus in the Western Hemisphere.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF197243; AAF22347.1; -.
DR INTERPRO; IPR000149; -.
DR PFAM; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
KW Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 >344 HEMAGGLUTININ.
FT NON_TER 344 344
SQ SEQUENCE 344 AA; 38322 MW; AD11B2D4EDA87F0A CRC64;

Query Match 40.6%; Score 82.5; DB 14; Length 344;
Best Local Similarity 60.0%; Pred. No. 0.011;
Matches 18; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

QY 6 QNTLKATGKKGPKYVKQNTLKATGKGV 35
|| | : | | | : | | | | | | | : |
Db 310 QNVNKVTYG-KCPKYIRQNTLKATGMNV 338

RESULT 14
Q906P0 PRELIMINARY; PRT; 344 AA.
AC Q906P0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HEMAGGLUTININ PRECURSOR (FRAGMENT).
GN HAI.
OS Equine influenza virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
OX NCBI_TaxID=11312;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A/EQ/SASKATOON/1/90;
RA Lai A.C.K.;
RT "Diverged evolution of equine-2 influenza (H3N8) virus in the Western Hemisphere.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF197243; AAF22347.1; -.
DR INTERPRO; IPR000149; -.
DR PFAM; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
KW Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 >344 HEMAGGLUTININ.
FT NON_TER 344 344
SQ SEQUENCE 344 AA; 38322 MW; AD11B2D4EDA87F0A CRC64;

Query Match 40.6%; Score 82.5; DB 14; Length 344;
Best Local Similarity 60.0%; Pred. No. 0.011;
Matches 18; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

QY 6 QNTLKATGKKGPKYVKQNTLKATGKGV 35
|| | : | | | : | | | | | | | : |
Db 310 QNVNKVTYG-KCPKYIRQNTLKATGMNV 338

RESULT 15
Q906N7 PRELIMINARY; PRT; 344 AA.
AC Q906N7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HEMAGGLUTININ PRECURSOR (FRAGMENT).
GN HAI.
OS Equine influenza virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
OX NCBI_TaxID=11312;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A/EQ/ARGENTINA/1/96;
RA Lai A.C.K.;
RT "Diverged evolution of equine-2 influenza (H3N8) virus in the Western Hemisphere.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF197246; AAF22350.1; -.
DR INTERPRO; IPR000149; -.
DR PFAM; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
KW Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 >344 HEMAGGLUTININ.
FT NON_TER 344 344
SQ SEQUENCE 344 AA; 38323 MW; 7321C177C01328E0 CRC64;

Query Match 40.6%; Score 82.5; DB 14; Length 344;
Best Local Similarity 60.0%; Pred. No. 0.011;
Matches 18; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

QY 6 QNTLKATGKKGPKYVKQNTLKATGKGV 35
|| | : | | | : | | | | | | | : |
Db 310 QNVNKVTYG-KCPKYIRQNTLKATGMNV 338

Search completed: April 14, 2001, 10:13:16
Job time: 221 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2001, 10:17:43 ; Search time 27.31 Seconds  
(without alignments)  
50.173 Million cell updates/sec

Title: US-09-362-731-4

Perfect score: 203

Sequence: 1 PKYVKQNTLKLATGKPKYKQNTLKLATGKKGVIIGIK 40

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	85.5	42.1	565	1 HEMA_IAHFO	P16995 Influenza a
2	85.5	42.1	565	1 HEMA_IAHNM	P16997 Influenza a
3	85.5	42.1	565	1 HEMA_IAHSA	P16999 Influenza a
4	83.5	41.1	566	1 HEMA_IAUDO	P19106 Influenza a
5	82.5	40.6	565	1 HEMA_IAHAL	P16994 Influenza a
6	82.5	40.6	565	1 HEMA_IAHK6	P19699 Influenza a
7	82.5	40.6	565	1 HEMA_IAHK7	P16996 Influenza a
8	82.5	40.6	565	1 HEMA_IAHSU	P08011 Influenza a
9	82.5	40.6	565	1 HEMA_IAHTE	P17001 Influenza a
10	81.5	40.1	328	1 HEMA_IAEN6	P04664 Influenza a
11	81.5	40.1	328	1 HEMA_IAEN7	P04663 Influenza a
12	81.5	40.1	550	1 HEMA_IABAN	P03441 Influenza a
13	81.5	40.1	550	1 HEMA_IADH1	P12582 Influenza a
14	81.5	40.1	550	1 HEMA_IADH2	P12583 Influenza a
15	81.5	40.1	550	1 HEMA_IADH3	P12584 Influenza a
16	81.5	40.1	550	1 HEMA_IADH4	P12585 Influenza a
17	81.5	40.1	550	1 HEMA_IADH6	P12587 Influenza a
18	81.5	40.1	550	1 HEMA_IADH7	P12588 Influenza a
19	81.5	40.1	550	1 HEMA_IADHM	P43259 Influenza a
20	81.5	40.1	550	1 HEMA_IADH2	P11133 Influenza a
21	81.5	40.1	565	1 HEMA_IAHMI	P16558 Influenza a
22	81.5	40.1	565	1 HEMA_IAHRO	P16998 Influenza a
23	81.5	40.1	565	1 HEMA_IAHTO	P17000 Influenza a
24	81.5	40.1	565	1 HEMA_IAHUR	P17002 Influenza a
25	81.5	40.1	566	1 HEMA_IAAIC	P03437 Influenza a
26	81.5	40.1	566	1 HEMA_IADU3	P03442 Influenza a
27	81.5	40.1	566	1 HEMA_IAEN7	P03440 Influenza a
28	81.5	40.1	566	1 HEMA_IAEN2	P03439 Influenza a
29	81.5	40.1	566	1 HEMA_IANT6	P03436 Influenza a
30	81.5	40.1	566	1 HEMA_IAZCO	P26139 Influenza a
31	81.5	40.1	567	1 HEMA_IAV17	P03435 Influenza a
32	80.5	39.7	566	1 HEMA_IAZUK	P26141 Influenza a
33	78.5	38.7	550	1 HEMA_IAME6	P12589 Influenza a

#### RESULT 1

ID	HEMA_IAHFO	STANDARD;	PRT;	565 AA.
AC	P16995; Q83993; Q83992;			
DT	01-AUG-1990 (Rel. 15, Created)			
DT	01-FEB-1991 (Rel. 17, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;			
DE	HEMAGGLUTININ HA2 CHAIN].			
GN	HA.			
OS	Influenza A virus (strain A/Equine/Fontainebleau/76) (Influenza A			
OS	virus (strain A/Equine/France/1/76)).			
OC	Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;			
OC	Influenza virus A and B group; Influenza A viruses;			
OC	Influenza A virus.			
OX	NCBI_TaxID=11399;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=89204899; Pubmed=2705299;			
RA	Kawaoka Y., Bean W.J., Webster R.G.;			
RT	"Evolution of the hemagglutinin of equine H3 influenza viruses.";			
RL	Virology 169:283-292(1989).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92114135; Pubmed=1731092;			
RA	Bean W.J., Schell M., Katz J., Kawaoka Y., Naeve C., Gorman O.,			
RA	Webster R.G.;			
RT	"Evolution of the H3 influenza virus hemagglutinin from human and			
RT	nonhuman hosts.";			
RL	J. Virol. 66:1129-1138(1992).			
CC	-!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO			
CC	CELL RECEPTORS AND FOR INITIATING INFECTION.			
CC	-!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS			
CC	(HA1 AND HA2) LINKED BY A DISULFIDE BOND.			
CC	-!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M24723; AAA43101.1; ALT SEQ.			
DR	EMBL; M73773; -; NOT_ANNOTATED_CDS.			
DR	PIR; F34064; HMIVE6.			
DR	HSSP; P03437; IHTM.			
DR	InterPro; IPR000149; -.			
DR	InterPro; IPR000386; -.			
DR	InterPro; IPR001364; -.			
DR	Pfam; PF00509; Hemagglutinin; 1.			
DR	PRINTS; PR00329; HEMAGGLUTN12.			
DR	PRINTS; PR00330; HEMAGGLUTN1.			
DR	PRINTS; PR00331; HEMAGGLUTN2.			
KW	Envelope protein; Hemagglutinin; Glycoprotein; Signal.			

P43257 influenza a  
P11134 influenza a  
P26135 influenza a  
P43258 influenza a  
P43260 influenza a  
P19694 influenza a  
P26134 influenza a  
P03449 influenza a  
P19696 influenza a  
P19698 influenza a  
P19695 influenza a

#### ALIGNMENTS

34 77.5 38.2 550 1 HEMA\_IADHK  
35 77.5 38.2 550 1 HEMA\_IADH3  
36 76.5 37.7 566 1 HEMA\_IADH2  
37 75.5 37.2 550 1 HEMA\_IADHL  
38 75.5 37.2 550 1 HEMA\_IAGHK  
39 75.5 37.2 564 1 HEMA\_IABUD  
40 75.5 37.2 566 1 HEMA\_IADA3  
41 75.5 37.2 566 1 HEMA\_IAMAO  
42 73.5 36.2 566 1 HEMA\_IAME1  
43 70.5 34.7 564 1 HEMA\_IADC2  
44 70.5 34.7 564 1 HEMA\_IAGRE  
45 69.5 34.2 564 1 HEMA\_IACKA

DR	PRINTS; PR00330; HEMAGGLUTN1.
DR	PRINTS; PR00331; HEMAGGLUTN2.
KW	Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT	SIGNAL 1 16
FT	CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT	CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT	CARBOHYD 23 23 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 37 37 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 180 180 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 498 498 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ	SEQUENCE 565 AA; 63749 MW; 35B4D05C374FB215 CRC64;

Query Match: 42.1%; Score 85.5; DB 1; Length 565;  
Best Local Similarity 63.3%; Pred. No. 0.0016;  
Matches 19; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY	6 QNTLKLTGKKPKYVKONTLKGKGV 35     :       :       :     :
DB	310 QNVNKVIYG-KCPKYIKONTLKGATGRNV 338     :       :       :     :

```

RESULT 3
HEMA_IAHSA
ID   HEMA_IAHSA      STANDARD;          PRT;    565 AA.
AC   P16999; Q84000; Q84001;
DT   01-AUG-1990 (Rel. 15, Created)
DD   01-FEB-1991 (Rel. 17, Last sequence update)
DI   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE   HEMAGGLUTININ HA2 CHAIN].
HA.
OS   Influenza A virus (strain A/Equine/Santiago/1/85).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza virus A and B group; Influenza A viruses;
OC   Influenza A virus.
OX   NCBI_TaxID=11414;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89204899; PubMed=2705299;
RA   Kawaoka Y., Bean W.J., Webster R.G.;
RT   "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL   Virology 169:283-292(1989).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC   CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC   (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
-----
CC   EMBL; M24725; AAA43110.1; ALT_SEQ.
DR   PIR; H34064; HMIVE8.
DR   HSSP; P03437; LHTM.
DR   InterPro; IPR000149; -.
DR   InterPro; IPR000386; -.
DR   InterPro; IPR001364; -.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   PRINTS; PR00330; HEMAGGLUTN1.
DR   PRINTS; PR00331; HEMAGGLUTN2.
KW   Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT   SIGNAL 1 16
FT   CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
```

```

FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63190 MW; 1B2F298F8AD800C2 CRC64;

Query Match 41.1%; Score 83.5; DB 1; Length 566;
Best Local Similarity 63.3%; Pred. No. 0.0028;
Matches 19; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

Qy 6 ONTKLATGKKGPKYVKONTLKLATGKKGV 35
Dy 311 QNVNKTIXG-ACPXYVKONTLKLATGMRNV 339

RESULT 5
HEMA_IAHAL
ID HEMA_TAHAL STANDARD; PRT; 565 AA.
AC P16994; Q83991; Q83990;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HAI CHAIN;
DE HEMAGGLUTININ HAZ CHAIN].
DE HA.
OS Influenza A virus (strain A/Equine/Algiers/72).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11393;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawaoka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial/
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M24721; AAA43100.1; ALT_SEQ.
DR PIR: D34064; HMIVE4.
DR HSP: P03437; LHTM.
DR InterPro: IPR000149; -
DR InterPro: IPR000386; -
DR InterPro: IPR001364; -
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR PRINTS: PR00330; HEMAGGLUTN1.
DR PRINTS: PR00331; HEMAGGLUTN2.
DR Envelope protein; Hemagglutinin; Glycoprotein; Signal.
KW FT SIGNAL 1 16
FT CHAIN 17 340 HEMAGGLUTININ HAI CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HAZ CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63831 MW; BA533050DC3F186B CRC64;

```



```
Db 310 QNVNKVTYG-KCPKYIRQNTLKLATGMNV 338

RESULT 8
HEMA_IAHSU
ID HEMA_IAHSU STANDARD; PRT; 565 AA.
AC Q08011;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].
GN HA.
OS Influenza A virus (strain A/Equine/Suffolk/89).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=45413;
RN RP SEQUENCE FROM N.A.
RX MEDLINE=93277383; PubMed=8503788;
RA Blinn M.M., Daly J.M., Chirnside E.D., Mumford J.A., Wood J.M.,
RA Richards C.M., Daniels R.S.;
RT "Genetic and antigenic analysis of an equine influenza H 3 isolate
RT from the 1989 epidemic.";
RL Arch. Virol. 130:33-44(1993).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X68437; CAA48482.1; -
CC HSSP; P03437; IHTM.
CC InterPro; IPR000149; -
CC InterPro; IPR000386; -
CC InterPro; IPR001364; -
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC PRINTS; PR00330; HEMAGGLUTN1.
CC PRINTS; PR00331; HEMAGGLUTN2.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN (BY SIMILARITY).
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN (BY SIMILARITY).
FT CARBOHYD 23 23 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 37 37 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63699 MW; C7A4E3B54B87D1A1 CRC64;

Query Match 40.6%; Score 82.5; DB 1; Length 565;
Best Local Similarity 60.0%; Pred. No. 0.0037;
Matches 18; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

QY 6 QNTLKLATGKKGPKYVQNTLKLATGKGV 35
|| : | | | | | | | | | |
Db 310 QNVNKVTYG-KCPKYIRQNTLKLATGMNV 338

RESULT 10
HEMA_IAEN6
ID HEMA_IAEN6 STANDARD; PRT; 328 AA.
AC P04664;

RESULT 9
HEMA_IAHTE
ID HEMA_IAHTE STANDARD; PRT; 565 AA.
AC P17001; Q84004; Q84005;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].
GN HA.
OS Influenza A virus (strain A/Equine/Tennessee/5/86).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11417;
RN RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=2705299;
RA Kawaoka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M24726; AAA43112.1; ALT_SEQ.
CC PIR; I34064; HMIVE9.
CC HSSP; P03437; IHTM.
CC InterPro; IPR000149; -
CC InterPro; IPR000386; -
CC InterPro; IPR001364; -
CC Pfam; PF00509; Hemagglutinin; 1.
CC PRINTS; PR00329; HEMAGGLUTN12.
CC PRINTS; PR00330; HEMAGGLUTN1.
CC PRINTS; PR00331; HEMAGGLUTN2.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 37 37 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63688 MW; IFE4485F0E/AC2C4 CRC64;

Query Match 40.6%; Score 82.5; DB 1; Length 565;
Best Local Similarity 60.0%; Pred. No. 0.0037;
Matches 18; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

QY 6 QNTLKLATGKKGPKYVQNTLKLATGKGV 35
|| : | | | | | | | | | |
Db 310 QNVNKVTYG-KCPKYIRQNTLKLATGMNV 338
```

```
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HEMAGGLUTININ [CONTAINS: HEMAGGLUTININ HA1 CHAIN] (FRAGMENT).
GN HA.
OS Influenza A virus (strain A/England/878/69).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81194918; PubMed=6164798;
RA Sleigh M.J., Both G.W., Underwood P.A., Bender V.J.;
RT "Antigenic drift in the hemagglutinin of the Hong Kong influenza
RT subtype: correlation of amino acid changes with alterations in viral
RT antigenicity.";
RL J. Virol. 37:845-853(1981).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K03335; AAA43184.1; -
DR InterPro; IPR001364; -
DR Pfam; PF00509; Hemagglutinin; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 328
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 63 63 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 328
FT SEQUENCE 328 AA; 36072 MW; 9C3A86B3ABD856FE CRC64;
SQ
Query Match 40.1%; Score 81.5; DB 1; Length 328;
Best Local Similarity 63.3%; Pred. No. 0.0029;
Matches 19; Conservative 2; Mismatches 8; Indels 1; Gaps 1;
QY 6 QNTLKATGKGGPKYKQNTLKATGKGV 35
DB 295 QNVNKITYG-ACPKYKQNTLKATGMRNV 323
|| : | ||||| : |
| : | ||||| : |
RESULT 11
HEMA_IAOU7
ID HEMA_IAOU7 STANDARD; PRT; 328 AA.
AC P04663;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HEMAGGLUTININ [CONTAINS: HEMAGGLUTININ HA1 CHAIN] (FRAGMENT).
GN HA.
OS Influenza A virus (strain A/Qu/7/70).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11456;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE=81194918; PubMed=6164798;
RA Sleigh M.J., Both G.W., Underwood P.A., Bender V.J.;
RT "Antigenic drift in the hemagglutinin of the Hong Kong influenza
RT subtype: correlation of amino acid changes with alterations in viral
RT antigenicity.";
RL J. Virol. 37:845-853(1981).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K03338; AAA43195.1; -
DR InterPro; IPR001364; -
DR Pfam; PF00509; Hemagglutinin; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 328
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 81 81 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 328
FT SEQUENCE 328 AA; 36059 MW; E30A962FD6DD805B CRC64;
SQ
Query Match 40.1%; Score 81.5; DB 1; Length 328;
Best Local Similarity 63.3%; Pred. No. 0.0029;
Matches 19; Conservative 2; Mismatches 8; Indels 1; Gaps 1;
QY 6 QNTLKATGKGGPKYKQNTLKATGKGV 35
DB 295 QNVNKITYG-ACPKYKQNTLKATGMRNV 323
|| : | ||||| : |
| : | ||||| : |
RESULT 12
HEMA_IABAN
ID HEMA_IABAN STANDARD; PRT; 550 AA.
AC P03441; Q83961; Q83962;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN] (FRAGMENT).
GN HA.
OS Influenza A virus (strain A/Bangkok/1/79).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11325;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82033259; PubMed=6169840;
RA Both G.W., Sleigh M.J.;
RT "Conservation and variation in the hemagglutinins of Hong Kong
RT subtype influenza viruses during antigenic drift.";
RL J. Virol. 39:663-672(1981).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
```



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: J02092; AAA43182.1; --  
 DR PIR: A04052; HMIIV6.  
 DR InterPro: IPR001364; --  
 DR Pfam: PF00509; Hemagglutinin; 1.  
 KW Envelope protein; Hemagglutinin; Glycoprotein.  
 FT NON\_TER 1 1  
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.  
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 550 AA; 61699 MW; FBD2EC200689CBE5 CRC64;

Query Match 40.1%; Score 81.5; DB 1; Length 550;  
 Best Local Similarity 63.3%; Pred. No. 0.0049; 8; Indels 1; Gaps 1;  
 Matches 19; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

QY 6 QNTLKATGKKGPKYKQNTLKATGKGV 35  
 || | : | ||||| : |  
 DB 295 QNVNKITYG-ACPXYKQNTLKATGMNV 323

RESULT 13  
 HEMA\_IADH1 STANDARD; PRT; 550 AA.  
 AC P12582; O84021; O84022;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;  
 DE HEMAGGLUTININ HA2 CHAIN] (FRAGMENT).  
 GN HA.  
 OS Influenza A virus (strain A/Duck/Hokkaido/5/77).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza virus A and B group; Influenza A viruses;  
 OC Influenza A virus.  
 OX NCBI\_TaxID=11357;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87265458; PubMed=2440178;  
 RA Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;  
 RT "Antigenic and genetic conservation of H3 influenza virus in wild  
 RT ducks";  
 RL Virology 159:109-119(1987).  
 CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.  
 CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS  
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
 CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR PIR: A27813; HMIIV77.  
 DR InterPro: IPR001364; --  
 DR Pfam: PF00509; Hemagglutinin; 1.  
 KW Envelope protein; Hemagglutinin; Glycoprotein.  
 FT NON\_TER 1 1  
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.  
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 550 AA; 61705 MW; 7E/ACFE716FC969A CRC64;

Query Match 40.1%; Score 81.5; DB 1; Length 550;  
 Best Local Similarity 63.3%; Pred. No. 0.0049; 8; Indels 1; Gaps 1;  
 Matches 19; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

QY 6 QNTLKATGKKGPKYKQNTLKATGKGV 35  
 || | : | ||||| : |  
 DB 295 QNVNKITYG-ACPXYKQNTLKATGMNV 323

RESULT 14  
 HEMA\_IADH2 STANDARD; PRT; 550 AA.  
 AC P12583; O84011;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;  
 DE HEMAGGLUTININ HA2 CHAIN] (FRAGMENT).  
 GN HA.  
 OS Influenza A virus (strain A/Duck/Hokkaido/8/80).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza virus A and B group; Influenza A viruses;  
 OC Influenza A virus.  
 OX NCBI\_TaxID=11358;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87265458; PubMed=2440178;  
 RA Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;  
 RT "Antigenic and genetic conservation of H3 influenza virus in wild  
 RT ducks";  
 RL Virology 159:109-119(1987).  
 CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.  
 CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS  
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
 CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: M16738; AAA43144.1; --  
 DR PIR: B27813; HMIIV80.  
 DR InterPro: IPR001364; --  
 DR Pfam: PF00509; Hemagglutinin; 1.  
 KW Envelope protein; Hemagglutinin; Glycoprotein.  
 FT NON\_TER 1 1  
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
 FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.  
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 137 137 K -> N (IN PIR DATA BANK).  
SQ SEQUENCE 550 AA; 61659 MW; A107023ACC9CC353 CRC64;

Query Match 40.1%; Score 81.5; DB 1; Length 550;  
Best Local Similarity 63.3%; Pred. No. 0.0049;  
Matches 19; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

QY 6 QNTLKLATGKKGPKYKQNTLKLATGKGV 35  
Db 295 QNVNKITYG-ACPKYKQNTLKLATGMRNV 323

## RESULT 15

HEMA\_IADH3  
ID HEMA\_IADH3 STANDARD; PRT; 550 AA.  
AC P12584; Q84012; Q89793;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;  
DE HEMAGGLUTININ HA2 CHAIN] (FRAGMENT).  
GN HA.  
OS Influenza A virus (strain A/Duck/Hokkaido/33/80).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses;  
OC Influenza A virus.  
OX NCBI\_TaxID=11359;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87265458; PubMed=2440178;  
RA Kida H., Kawaka Y., Naeye C.W., Webster R.G.;  
RT "Antigenic and genetic conservation of H3 influenza virus in wild  
RT ducks.";  
RL Virology 159:109-119(1987).  
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO  
CC CELL RECEPTORS AND FOR INITIATING INFECTION.  
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS  
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.  
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M16739; AAA43145.1; -;  
DR PIR; C27813; HMIV33.  
DR InterPro; IPR001364; -;  
DR Pfam; PF00509; Hemagglutinin; 1.  
KW Envelope protein; Hemagglutinin; Glycoprotein.  
FT NON\_TER 1 1  
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.  
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.  
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 550 AA; 61577 MW; 6C30BF67CFDCB7DE CRC64;

Query Match 40.1%; Score 81.5; DB 1; Length 550;  
Best Local Similarity 63.3%; Pred. No. 0.0049;  
Matches 19; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

QY 6 QNTLKLATGKKGPKYKQNTLKLATGKGV 35

Db 295 QNVNKITYG-ACPKYKQNTLKLATGMRNV 323

Search completed: April 14, 2001, 10:17:44  
Job time: 346 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2001, 10:11:49 ; Search time 46.83 Seconds  
(without alignments)  
58.700 Million cell updates/sec

Title: US-09-362-731-4  
Perfect score: 203  
Sequence: 1 PRYVKQNTLKLATGKGPYKQNTLKLATGKGVIGIK 40

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues  
Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_67:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	85.5	42.1	565	1 HMIWE5	hemagglutinin prec
2	85.5	42.1	565	1 HMIWE6	hemagglutinin prec
3	85.5	42.1	565	1 HMIWE8	hemagglutinin prec
4	82.5	40.6	565	1 HMIWE4	hemagglutinin prec
5	82.5	40.6	565	1 HMIWE9	hemagglutinin prec
6	82.5	40.6	565	1 HMIWE7	hemagglutinin prec
7	82.5	40.6	565	1 HMIWE6	hemagglutinin prec
8	82.5	40.6	565	2 S33703	hemagglutinin - in
9	81.5	40.1	550	1 HMIW53	hemagglutinin prec
10	81.5	40.1	550	1 HMIW77	hemagglutinin prec
11	81.5	40.1	550	1 HMIW80	hemagglutinin prec
12	81.5	40.1	550	1 HMIW33	hemagglutinin prec
13	81.5	40.1	550	1 HMIW89	hemagglutinin prec
14	81.5	40.1	550	1 HMIW98	hemagglutinin prec
15	81.5	40.1	550	1 HMIW15	hemagglutinin prec
16	81.5	40.1	550	2 JQ1156	hemagglutinin prec
17	81.5	40.1	565	1 HMIWE1	hemagglutinin prec
18	81.5	40.1	565	1 HMIWE2	hemagglutinin prec
19	81.5	40.1	565	1 HMIWE3	hemagglutinin prec
20	81.5	40.1	565	1 HMIWE7	hemagglutinin prec
21	81.5	40.1	566	1 HMIW1H	hemagglutinin prec
22	81.5	40.1	566	1 HMIW1H	hemagglutinin prec
23	81.5	40.1	566	1 HMIW1H	hemagglutinin prec
24	81.5	40.1	566	1 HMIW6	hemagglutinin prec
25	81.5	40.1	566	1 HMIW1H	hemagglutinin prec
26	81.5	40.1	567	2 S52188	hemagglutinin - in
27	78.5	38.7	347	2 HMIW86	hemagglutinin prec
28	78.5	38.7	550	1 HMIW52	hemagglutinin prec
29	77.5	38.2	550	1 HMIW52	hemagglutinin prec

30	77.5	38.2	550	2 JQ1153	hemagglutinin prec
31	75.5	37.2	330	2 JQ2374	hemagglutinin - in
32	75.5	37.2	330	2 JQ2375	hemagglutinin - in
33	75.5	37.2	331	2 JQ2377	hemagglutinin - in
34	75.5	37.2	331	2 JQ2378	hemagglutinin - in
35	75.5	37.2	347	2 S52173	hemagglutinin - in
36	75.5	37.2	347	2 S52174	hemagglutinin - in
37	75.5	37.2	347	2 S52175	hemagglutinin - in
38	75.5	37.2	347	2 S52176	hemagglutinin - in
39	75.5	37.2	347	2 S52178	hemagglutinin - in
40	75.5	37.2	347	2 S52180	hemagglutinin - in
41	75.5	37.2	347	2 S52182	hemagglutinin - in
42	75.5	37.2	347	2 S52183	hemagglutinin - in
43	75.5	37.2	347	2 S52185	hemagglutinin - in
44	75.5	37.2	347	2 S52186	hemagglutinin - in
45	75.5	37.2	347	2 S52199	hemagglutinin - in

ALIGNMENTS

RESULT 1

HMIWE5

hemagglutinin precursor - influenza A virus (strain A/equine/New Market/76[H3N8])  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C:Date: 30-Sep-1990 #sequence\_revision 30-Sep-1990 #text\_change 16-Jul-1999  
C:Accession: E34064  
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.  
Virology 169, 283-292, 1989  
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.  
A:Reference number: A34064; MUID:89204899  
A:Accession: E34064  
A:Molecule type: genomic RNA  
A:Residues: 1-565 <KAW>  
A:Cross-references: GB:M24722; GB:J04336; NID:g324010; PIDN:AAA43107.1; PID:g324011  
C:Genetics:  
C:Superfamily: influenza virus hemagglutinin  
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-344/Product: hemagglutinin HA1 #status predicted <HAL>  
F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>  
F:535-551/Domain: transmembrane #status predicted <TM1>  
F:23-37,53,78,180,300,498/Binding site: Carbohydrate (Asn) (covalent) #status predicted  
F:29-481,67-292,79-91,154-488,296-320/Bisulfide bonds: #status predicted  
F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 42.1%; Score 85.5; DB 1; Length 565;  
Best Local Similarity 63.3%; Pred No. 0.0079; 7; Indels 1; Gaps 1;  
Matches 19; Conservative 3; Mismatches 7;

QY 6 QNTLKLATGKGPYKQNTLKLATGKGV 35  
||| : | : ||| : ||| : ||| : |  
Db 310 QNVKVTYG-KCPKVIKQNTLKLATGMRNV 338

RESULT 2

HMIWE6

hemagglutinin precursor - influenza A virus (strain A/equine/Fontainebleau/76[H3N8])  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C:Date: 30-Sep-1990 #sequence\_revision 30-Sep-1990 #text\_change 16-Jul-1999  
C:Accession: F34064  
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.  
Virology 169, 283-292, 1989  
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.  
A:Reference number: A34064; MUID:89204899  
A:Accession: F34064  
A:Molecule type: genomic RNA  
A:Residues: 1-565 <KAW>  
A:Cross-references: GB:M24723; GB:J04336; NID:g323998; PIDN:AAA43101.1; PID:g323999

A:Cross-references: GB:M24721; GB:J04336; NID:G323996; PIDN:AAA43100.1; PID:G323997

C:Genetics:

A:Map position: segment 4

C:Superfamily: influenza virus hemagglutinin

C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>

F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>

F:535-551/Domain: transmembrane #status predicted <TM1>

F:23-37,53,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted

F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 40.6%; Score 82.5; DB 1; Length 565;  
Best Local Similarity 63.3%; Pred. No. 0.018; Mismatches 19; Conservative 3; Indels 1; Gaps 1;

Oy 6 QNTLKLATGGKPKYVKQNTLKLATGKKGV 35  
|| : | |||::||||| : |  
Db 310 QNVNKITYG-KCPKYVQSTLKLATGMNV 338

RESULT 5  
HMIVS9  
hemagglutinin precursor - influenza A virus (strain A/equine/Tennessee/5/85[H3N8])  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: Influenza A virus  
C:Date: 30-Sep-1990 #sequence\_revision 30-Sep-1990 #text\_change 16-Jul-1999  
C:Accession: J34064  
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.  
Virology 169, 283-292, 1989  
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.  
A:Reference number: A34064; PMID:89204899  
A:Accession: J34064  
A:Molecule type: genomic RNA  
A:Residues: 1-565 <KAW>  
A:Cross-references: GB:M24726; GB:J04336; NID:G324020; PIDN:AAA43112.1; PID:G324021

C:Genetics:

A:Map position: segment 4

C:Superfamily: influenza virus hemagglutinin

C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>

F:345-565/Product: hemagglutinin HA2 #status predicted <HA2>

F:535-551/Domain: transmembrane #status predicted <TM1>

F:18,23,37,53,68,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted

F:554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 40.6%; Score 82.5; DB 1; Length 565;  
Best Local Similarity 60.0%; Pred. No. 0.018; Mismatches 18; Conservative 4; Indels 1; Gaps 1;

Oy 6 QNTLKLATGGKPKYVKQNTLKLATGKKGV 35  
|| : | |||::||||| : |  
Db 310 QNVNKITYG-KCPKYIRQNTLKLATGMNV 338

RESULT 6  
HMIPT  
hemagglutinin precursor - influenza A virus (strain A/equine/Kentucky/2/86[H3N8])  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: Influenza A virus  
C:Date: 30-Sep-1990 #sequence\_revision 30-Sep-1990 #text\_change 16-Jul-1999  
C:Accession: A34065  
R:Kawaoka, Y.; Bean, W.J.; Webster, R.G.  
Virology 169, 283-292, 1989  
A:Title: Evolution of the hemagglutinin of equine H3 influenza viruses.  
A:Reference number: A34064; PMID:89204899  
A:Accession: A34065  
A:Molecule type: genomic RNA

```
A;Residues: 1-565 <KAW>
A;Cross-references: GB:M24727; GB:J04336; NID:g324000; PIDN:AAA43102.1; PID:g324001
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F;345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F;535-551/Domain: transmembrane #status predicted <TM1>
F;18,23,37,53,68,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F;554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 40.6%; Score 82.5; DB 1; Length 565;
Best Local Similarity 60.0%; Pred. No. 0.018;
Matches 18; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

QY 6 QNTLKLATGKKPKYKQNTLKLATGKKGV 35
|| | | | | | | | | | | | | | | | | |
Db 310 QNVNKVTYG-KCPKYIRQNTLKLATGMRNV 338

RESULT 7
HMI7EE
hemagglutinin precursor - influenza A virus (strain A/equine/Kentucky/1/87[H3N8])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C;Accession: B34065
R;Kawaoka, Y.; Bean, W.J.; Webster, R.G.
Virology 169, 283-292, 1989
A;Title: Evolution of the hemagglutinin of equine H3 influenza viruses.
A;Reference number: A34064; MUID:89204899
A;Accession: B34065
A;Molecule type: genomic RNA
A;Residues: 1-565 <KAW>
A;Cross-references: GB:M24728; GB:J04336; NID:g324002; PIDN:AAA43103.1; PID:g324003
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F;345-565/Product: hemagglutinin HA2 #status predicted <HA2>
F;535-551/Domain: transmembrane #status predicted <TM1>
F;18,23,37,53,68,78,180,300,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;29-481,67-292,79-91,154-488,296-320/Disulfide bonds: #status predicted
F;554,561,564/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 40.6%; Score 82.5; DB 1; Length 565;
Best Local Similarity 60.0%; Pred. No. 0.018;
Matches 18; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

QY 6 QNTLKLATGKKPKYKQNTLKLATGKKGV 35
|| | | | | | | | | | | | | | | | | |
Db 310 QNVNKVTYG-KCPKYIRQNTLKLATGMRNV 338

RESULT 8
HMI777
hemagglutinin - influenza A virus H3N8
C;Species: influenza A virus H3N8, equine influenza virus
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Sep-1999
C;Accession: S33703
R;Binns, M.M.; Daly, J.M.; Chirnside, E.D.; Mumford, J.A.; Wood, J.M.; Richards, C.M.; D
Arch. Virol. 130, 33-43, 1993
A;Title: Genetic and antigenic analysis of an equine influenza H3 isolate from the 1989
A;Reference number: S33703; MUID:93277383
A;Accession: S33703
A;Status: preliminary
```

```
A;Molecule type: genomic RNA
A;Residues: 1-565 <BIN>
A;Cross-references: EMBL:X68437; NID:g312668; PIDN:CAA48482.1; PID:g312669
A;Note: the authors translated the codon ACC for residue 403 as Arg
C;Superfamily: influenza virus hemagglutinin

Query Match 40.6%; Score 82.5; DB 2; Length 565;
Best Local Similarity 60.0%; Pred. No. 0.018;
Matches 18; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

QY 6 QNTLKLATGKKPKYKQNTLKLATGKKGV 35
|| | | | | | | | | | | | | | | | | |
Db 310 QNVNKVTYG-KCPKYIRQNTLKLATGMRNV 338

RESULT 9
HMI753
hemagglutinin precursor - influenza A virus (strain A/swine/81/78) (fragment)
C;Species: influenza A virus
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C;Accession: B29971
R;Kida, H.; Shortridge, K.F.; Webster, R.G.
Virology 162, 160-166, 1988
A;Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China
A;Reference number: A94370; MUID:88101364
A;Accession: B29971
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M19057; NID:g324210
A;Note: the sequence in GenBank entry FLAHAPB, release 106, (PID:g324211) differs fro
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;300-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 40.1%; Score 81.5; DB 1; Length 550;
Best Local Similarity 63.3%; Pred. No. 0.023;
Matches 19; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

QY 6 QNTLKLATGKKPKYKQNTLKLATGKKGV 35
|| | | | | | | | | | | | | | | | | |
Db 295 QNVNKVTYG-ACPKYKQNTLKLATGMRNV 323

RESULT 10
HMI777
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/5/77) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C;Accession: A27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458
A;Accession: A27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16737; NID:g324081; PIDN:AAA43143.1; PID:g324082
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
```

Query Match 40.1%; Score 81.5; DB 1; Length 550;  
Best Local Similarity 63.3%; Pred. NO. 0.023;  
Matches 19: Conservative 2; Mismatches 8; Indels 1;  
Gaps 1;

QY 6 QNTLKLATGKKGPKYVVKQNTLKLATGKKGV 35  
 || : | ||||| : |  
 pb 295 QNVNKITYG-ACPKYVKQNTLKLATGMNV 323

```

RESULT      13
HMI489      hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/7/82) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: D27813
R:Kida, H.; Kawaoaka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; MUID:87265458
A:Accession: D27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:161740; NID:g324087; PID:AAA43146.1; PID:g324088
C:Genetics:
C:Map position: segment 4

```

Query Match 40.1%; Score 81.5; DB 1; Length 550;  
Best Local Similarity 63.3%; Pred. No. 0.023;  
Matches 19; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

QY 6 QNTLKATGKKGPKYVKQNTLKLATGKGV 35  
DB 295 QNVNKITYG-ACPKYVKQNTLKLATGMRNV 323

RESULT 14  
HMIv98  
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/9/85) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2  
C:Species: influenza A virus  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 18-Sep-1998  
C:Accession: F27813  
R:Kida, H.; Kawaoaka, Y.; Naeve, C.W.; Webster, R.G.  
Virology 159, 109-119, 1987  
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.  
A:Reference number: A94363; MUID:87265458  
A:Accession: F27813  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <KID>  
A:Cross-references: GB:M16742; NID:g324091  
C:Genetics:  
A:Map position: segment 4

F:320-550/Product: hemagglutinin HA2 #status predicted <HA2>  
F:530-536/Domains: transmembrane #status predicted <TM1>  
K:keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted  
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 40.1%; Score 81.5; DB 1; Length 550;

Best Local Similarity 63.3%; Pred. No. 0.023; Mismatches 2; Indels 1; Gaps 1;

Matches 19; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

QY 6 QNTLKLATGKGGPKYVQNTLKLATGKGGV 35

|| | : | ||||| : |

Db 295 QNVNKITYG-ACPKYVQNTLKLATGMRNV 323

#### RESULT 15

HMIV15

hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/10/85) (fragment)  
N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: Influenza A virus

C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 16-Jul-1999

C:Accession: G27813

R:Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.

Virology 159, 109-119, 1987

A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.

A:Reference number: A94363; MUID:87265458

A:Accession: G27813

A:Molecule type: genomic RNA

A:Residues: 1-550 <KID>

A:Cross-references: GB:M16743; NID:G324093; PIDN:AAA3149.1; PID:G324094

C:Genetics:

A:Map position: segment 4

C:Superfamily: influenza virus hemagglutinin

C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F:520-536/Domain: transmembrane #status predicted <TM1>

F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 40.1%; Score 81.5; DB 1; Length 550;

Best Local Similarity 63.3%; Pred. No. 0.023;

Matches 19; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

QY 6 QNTLKLATGKGGPKYVQNTLKLATGKGGV 35

|| | : | ||||| : |

Db 295 QNVNKITYG-ACPKYVQNTLKLATGMRNV 323

Search completed: April 14, 2001, 10:11:50  
Job time: 136 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2001, 10:13:09 : Search time 77.31 Seconds  
(without alignments)  
207.703 Million cell updates/sec

Title: US-09-362-731-3

Perfect score: 793

Sequence: 1 DQYKANSKFIGITELGGQY.....FGGCHGSEPCIIHRGKPFESR 137

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_15:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_unclassified:\*
- 13: sp\_vertebrate:\*
- 14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	104	13.1	170	10	O9LIX2	O9LIX2 oryza sativ
2	101	12.7	1571	11	O54978	O54978 mus musculu
3	94.5	11.9	3680	5	O9VR08	O9VR08 drosophila
4	91	11.5	341	5	O9VY49	O9VY49 drosophila
5	89.5	11.3	270	5	P90606	P90606 trypanosoma
6	88	11.1	2570	4	O9NY15	O9NY15 homo sapien
7	86	10.8	488	11	O9UJRO	O9UJRO mus musculu
8	85	10.7	169	4	O14564	O14564 homo sapien
9	85	10.7	232	5	O9U2R9	O9U2R9 caenorhabdi
10	85	10.7	2212	4	O93072	O93072 homo sapien
11	84	10.6	230	11	O64507	O64507 mus musculu
12	84	10.6	554	4	O9UH16	O9UH16 homo sapien
13	82.5	10.4	1378	11	O61138	O61138 mus musculu
14	82	10.3	102	1	O9YAA8	O9YAA8 acroporum p
15	81.5	10.3	1584	5	O93791	O93791 caenorhabdi
16	81	10.2	441	5	O9W5X1	O9W5X1 drosophila
17	81	10.2	846	2	O05442	O05442 mycobacteri
18	80.5	10.2	107	5	O9NG19	O9NG19 crassostrea
19	80	10.1	361	10	Q42640	Q42640 brassica ol

20	80	10.1	1523	11	O88280	O88280 rattus norv
21	79.5	10.0	1387	4	Q9UPN8	Q9UPN8 homo sapien
22	79.5	10.0	1833	11	O08999	O08999 mus musculu
23	79.5	10.0	2026	4	O00468	O00468 homo sapien
24	79	10.0	1637	6	O9XSV8	O9XSV8 bos taurus
25	78.5	9.9	328	5	O9VLB4	O9VLB4 drosophila
26	78.5	9.9	1664	5	O9TVQ2	O9TVQ2 caenorhabdi
27	78	9.8	809	5	Q9U3W7	Q9U3W7 drosophila
28	78	9.8	1114	11	O9JKW7	O9JKW7 mus musculu
29	78	9.8	1742	5	O9VD94	O9VD94 drosophila
30	78	9.8	1821	4	O14767	O14767 homo sapien
31	77.5	9.8	471	5	O9UAY0	O9UAY0 caenorhabdi
32	77.5	9.8	752	13	O42374	O42374 brachydanio
33	77	9.7	988	6	O97867	O97867 sus scrofa
34	77	9.7	1698	5	O94438	O94438 chironomus
35	76.5	9.6	3871	5	O20911	O20911 caenorhabdi
36	76	9.6	74	5	O9VIA1	O9VIA1 drosophila
37	76	9.6	242	2	O9LON6	O9LON6 streptomyce
38	76	9.6	261	11	O60493	O60493 cavia porce
39	76	9.6	313	10	O9SYZ5	O9SYZ5 arabidopsis
40	76	9.6	419	10	O42638	O42638 brassica ol
41	76	9.6	1487	10	O9MA08	O9MA08 arabidopsis
42	75.5	9.5	373	10	Q9SAW2	Q9SAW2 arabidopsis
43	75.5	9.5	646	10	O9SH64	O9SH64 arabidopsis
44	75	9.5	135	5	Q9TZ22	Q9TZ22 euroglyphus
45	75	9.5	145	5	O96430	O96430 euroglyphus

#### ALIGNMENTS

RESULT 1

O9LIX2 PRELIMINARY; PRT; 170 AA.

AC O9LIX2;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE RICE EST C27893 CORRESPONDS TO A REGION OF THE PREDICATED GENE.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hsiao Y.C., Chow T., Chen C., Wu H., Chu M., Chao Y., Liu S.;  
RT "Oryza sativa PAC P0699E04 genomics sequence, complete sequence."  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP001111; BAA90497.1; -  
SQ SEQUENCE 170 AA; 16813 MW; A9F1CD6FD3318A01 CRC64;

Query Match 13.1%; Score 104; DB 10; Length 170;

Best Local Similarity 34.7%; Pred. No. 0.00054;

Matches 33; Conservative 2; Mismatches 28; Indels 32; Gaps 5;

Qy 29 GTTESSCHGSEPCIIHRGKPGGCHGSEPCIIHRGKPFSSCHGSEPCIIHRGKPGGCH 88

Db 97 GAHKLHSHGG-----HGGYGYGG-----HHGGLFGGHHG-----HHGGLFGGHH 135

Qy 89 GSEPCIIHRGKPFSSCHGSEPCIIHRGKPGGCHG 123

Db 136 GH-----HGGGLFGGHHG-----HHGGLFGGHHG 159

RESULT 2

O54978 PRELIMINARY; PRT; 1571 AA.

ID O54978

AC O54978;

DT 01-JUN-1998 (TREMBLrel. 06, Created)

DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)

DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE ZINC FINGER PROTEIN.

[illegible]

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Anandides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,  
 RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abrah J.F., Agbayani A., An H.-J., Andrews-pfannkuch C., Baldwin D.,  
 RA Ballaw R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottiler P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Hostin N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Houtin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2193(2000).  
 DR EMBL; AE003495; AAF48356.1;  
 DR FLYBASE; FBgn0030551; CG11674.  
 DR INTERPRO; IPR001304;  
 DR PFAM; PF01683; EB; 1;  
 DR INTERPRO; IPR002899;  
 DR PROSITE; PS00615; C-TYPE-LECTIN.1; UNKNOWNW.1.  
 SQ SEQUENCE 341 AA; 37711 MW; 70CDAAB4053B7B24 CRC64;

Query Match 11.5%; Score 91; DB 5; Length 341;  
 Best Local Similarity 25.9%; Pred. No. 0.028;  
 Matches 36; Conservative 12; Mismatches 45; Indels 46; Gaps 10;

QY 31 TELSCHGSE-----PCIIHRGKPGGCHGSEPCIIHRGKPFSSCHGSEPCI- 77  
 Db 86 TRWQCHCSEGHVSDDRRCLFAVPVG---GSCEFOQOC--QRADRFSSCIGNQ-CLC 139  
 QY 78 -----IHRG-----KPFSGCHGSEPCIIHR-----GKPFSSCHGSEPCIIHR 114  
 Db 140 LNQFEPHEGRCLSVLOSSCLEDDKCGSC-CASICLTTKRCGCKNFVHNHNTKCI--K 196  
 QY 115 GRPFGG-CHGSEPCIIHRG 132  
 Db 197 GSAYGDTCEHSSPKLNLG 215

RESULT 5  
 P90606 ID P90606 PRELIMINARY; PRT; 270 AA.

AC P90606;  
 DT 01-MAY-1997 (TremBLrel. 03, Created)  
 DT 01-MAY-1997 (TremBLrel. 03, Last sequence update)  
 DT 01-JUN-2000 (TremBLrel. 14, Last annotation update)  
 DE NUCLEIC ACID BINDING PROTEIN.  
 OS Trypanosoma equiperdum.

OC Eukaryota; Eulenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5694;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strand A.D., Eisen H.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U87108; AAB47542.1; -;  
 DR INTERPRO; IPR001878; -;  
 DR PFAM; PF00098; zf-CCHC; 7.  
 DR PRINTS; PR00939; C2HCZNFINGER.  
 SQ SEQUENCE 270 AA; 28453 MW; 50D2C559FD9F7234 CRC64;

Query Match 11.3%; Score 89.5; DB 5; Length 270;  
 Best Local Similarity 33.9%; Pred. No. 0.032;  
 Matches 38; Conservative 4; Mismatches 43; Indels 27; Gaps 9;

QY 37 HGSEPCIIHRG-KPFGGCHGSEPCIIHRGKPFSSCHGSEPCIIHRGKPFGGC--HGSEPC 93  
 Db 53 HLSRDCPSNRGTAPMGGRACYNC---GQP---GHFSRECPNMRGMPGGMGGGRAC 105  
 QY 94 IHRGKPFSSC---HGSEPCIIHRGKPFGGC--HGSEPCIIHRGKPF--FSR 137  
 Db 106 Y-----NCVOPGHFSRECPNMRGMPGGMGGGRAC-YHCGQPGHFSR 148

RESULT 6  
 Q9NY15 ID Q9NY15 PRELIMINARY; PRT; 2570 AA.  
 AC Q9NY15;  
 DT 01-OCT-2000 (TremBLrel. 15, Created)  
 DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TremBLrel. 15, Last annotation update)  
 DE STABILIN-1.  
 GN STAB1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Politz O., Guillot P., Gratchev A., Schledzewski K., Birk R.,  
 RA Hakly N., Tebbe B., Orfanos C.E., Goerdts S.;  
 RT "Stabilin-1: an endothelial-macrophage member of the fasciclin domain  
 containing protein family associated with angiogenesis.";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ275213; CAB61827.1; -;  
 SQ SEQUENCE 2570 AA; 275346 MW; 3123FABD7C8E2BF8 CRC64;

Query Match 11.1%; Score 88; DB 4; Length 2570;  
 Best Local Similarity 28.6%; Pred. No. 0.49;  
 Matches 34; Conservative 7; Mismatches 40; Indels 38; Gaps 10;

QY 35 SCHGSEPCIIHRGKPFGGCH-----GSEPCIIHRGKPFSS--CHGSEPCIIHRGK 82  
 Db 858 SCTPSPNC-SHPDR--GGCSENAECVPSLGHCHTCHKGWSGDRVCVAIDCELDVG- 913  
 QY 83 PFGGCH-----GSEPCIIHRGKPFSS-----SCHGSEPCIIHRGKPFGGCHGSEPC 127  
 Db 914 --GGCHTDALCSYVGPQSRCTKLG--FAGDGYQCSPIDPC--RAGN--GGCHGLATC 964

RESULT 7  
 Q9JJR0 ID Q9JJR0 PRELIMINARY; PRT; 488 AA.

AC Q9JJR0;  
 DT 01-OCT-2000 (TremBLrel. 15, Created)  
 DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TremBLrel. 15, Last annotation update)  
 DE ZINC FINGER PROTEIN 92.  
 GN ZFP92.  
 OS Mus musculus (Mouse).

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RA Mallon A.M., Platzer M., Bates R., Gloeckner G., Botcherby M.,
RA Nordstiek G., Strivens M.A., Kloschis P., Dangel A., Cunningham D.,
RA Straw R., Weston P., Hunter C., Gilbert M., Fernando S., Goodall K.,
RA Kerry G., Greystrong J.S., Clark D., Goerdes M., Blechschmidt K., B.,
RA Rump A., Hinzmann B., Mundy C.R., Miller W., Poustka A., Herman G.E.,
RA Rhodes M., Denny P., Rosenthal A., Brown S.D.M.;
RT "Comparative genome sequence analysis of the Bpa/Str region in mouse
RT and man.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL049866; CAB88276.1; -.
SQ SEQUENCE 488 AA; 55975 MW; E163773E0903D68C CRC64;

Query Match 10.8%; Score 86; DB 11; Length 488;
Best Local Similarity 33.6%; Pred. No. 0.14;
Matches 49; Conservative 10; Mismatches 39; Indels 48; Gaps 15;

QY 14 TELGGQYIKANSKFTIGITELSSCH--GSEPCIIHR-GKPPGG---CHGSEPCIIHRG-KP 66
Db ||| ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
256 TEGC---KAFSRSSNLIEHQRIHSGQPKYCKEGKAFKGVSLIHQ---LIHRGDKP 308

QY 67 PSSCHGSEPCIIHRGKPGGCHG-SEPCIIHRG-KPF--SSCHGSEPCIIHRGKPPGG-- 120
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
309 F-TCH-----EYKAFRGLSLGSLQHVRHGXKPYECSEC-----GRAFGRR 350

QY 121 -----CHGSEPCIIH--RGKPPFR 137
Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
351 NLFKHQVHGVVR-LQHTRGKGFOR 375

RESULT 8
ID Q14564 PRELIMINARY; PRT; 169 AA.
AC Q14564;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HIGH-SULPHUR KERATIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTES;
RA Drabant B., Doenecke D.;
RL Submitted (DEC-1991) to the EMBL/GenBank/DBJ databases.
DR EMBL; X63755; CAA45283.1; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR001007; -.
DR INTERPRO; IPR001138; -.
DR INTERPRO; IPR001368; -.
DR PROSITE; PS00463; ZN2_CV6_FUNGAL_1; UNKNOWN_1.
DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR PROSITE; PS01208; VWFC; UNKNOWN_2.
KW Keratin.
SQ SEQUENCE 169 AA; 16216 MW; A39206EB0B49D4BE CRC64;

Query Match 10.7%; Score 85; DB 4; Length 169;
Best Local Similarity 27.9%; Pred. No. 0.06;
Matches 34; Conservative 9; Mismatches 53; Indels 26; Gaps 8;

QY *12 GITELGQYIKANSKFTIGITELSSCHGSEPCIIHRGKPPGGCHGSEPCIIHRGKPP--SSC 70
Db ||| ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
61 GCGSGGGS--KGGCGSGCGSQSCC---KPCCCS-----SGC-GSSCCQCCOCKPYCSQC 109
```

```
QY 71 HGSEPCIIHRGKPPFGGCHGS--EPCIIHRGKPPFSCHGS--EPCIIHRGKPPFGGCHGSEP 126
Db ||| ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
110 SCCCKPCSSSGRSCCQSSCKKPCSSSGSGSCCQSCCKPC-----CSQSRC 159

QY 127 CI 128
Db |||
160 CV 161

RESULT 9
ID Q9U2R9 PRELIMINARY; PRT; 232 AA.
AC Q9U2R9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE Y116A8C.41 PROTEIN.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AL117204; CAB55158.1; -.
DR INTERPRO; IPR000168; -.
SQ SEQUENCE 232 AA; 27290 MW; 592B47AB075E9768 CRC64;

Query Match 10.7%; Score 85; DB 5; Length 232;
Best Local Similarity 35.6%; Pred. No. 0.083;
Matches 36; Conservative 6; Mismatches 35; Indels 24; Gaps 10;

QY 45 HRGKPP-GGCHGSEPCIIHRGKPP-SSCHGSEPCIIHRGKPP-GGCHGSEPCII--IHRGK 99
Db ||| ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
81 HRGETLGRTHRGE---THRGETFRGETHRGE---THRGETFRGETHRGETFCGETHRGE 134

QY 100 PF-SSCHGSEPCIIHRGKPP-----GGCHGSEPCIIHRGK 133
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
135 TFRSETHRGE---THRGETFRGETFRGEAHRGE---AHRGE 169

RESULT 10
ID Q93072 PRELIMINARY; PRT; 2212 AA.
AC Q93072;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE MYELOBLAST KIAA0246 PROTEIN (FRAGMENT).
OS KIAA0246.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW;
RX MEDLINE=97191544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,
RA Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.";
```

Db 69 SSCGGCGSC-----GSSCCG-C- -KPVCCCVPCVCSGCCGCKPCCQSSCCK 119

QY 83 P--FGCGHSGEPCIHRGKPF---SSCHGSEPCIHRG--KP--FGGCHGSEPC 127  
| | | | | : | | : | | | | |  
Db 119 PCCSSGC-GSSCCQSSCCKPCCQSSC--CKPCCQSSCCKPCCSSGC-GSSCC 168

RESULT 12

ID Q9UH16 PRELIMINARY; PRT; 554 AA.

AC Q9UH16;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE D16ZHL14.1 (FIBULIN 1) (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Lloyd D.;

RL Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.

RR EMBL; Z98047; CAB62995.1; -.

DR HSP; P35555; IEMN.

DR INTERPRO: IPR000020; -.

DR INTERPRO: IPR00152; -.

DR INTERPRO: IPR000561; -.

DR INTERPRO: IPR001881; -.

DR PFAM; PF00008; EGF; 6.

DR PFAM; PF01821; ANATO; 3.

DR PROSITE; PS00010; ASX\_HYDROXYL; 4.

DR PROSITE; PS01177; ANAPHYLATOXIN\_1; 3.

DR PROSITE; PS01186; EGF\_2; 3.

FT NON\_TER

SQ SEQUENCE 554 AA; 59767 MW; FE285184599A2982 CRC64;

Query Match 10.6%; Score 84; DB 4; Length 554;  
Best Local Similarity 30.9%; Pred. No. 0.26;  
Matches 34; Conservative 8; Mismatches 44; Indels 24; Gaps

QY 29 GIWELSSCHGSEPCIHRGKPFGGCHGSEPCIHRGKPF-----SCHGSEPCIHRGKP 83  
| | | | | : | | | | | : | | : | | : | |  
Db 414 GINECSSPCSQC-----ANYGSYCYCRGYQLSDVDGVTCESIDECAL----P 461

QY 84 FGCGHSGEPCIHRGKPFSSCHGSEPCIHRGKPFPG-GCHGSEPCI--IH 130  
| | | | | : | | | | | : | | : | | : | |  
Db 462 TGGHICSYRCINIPGSFQSC---PSSGYRLAPGRNCQGIDECAVTGIH 507

RESULT 13

Q61138 PRELIMINARY; PRT; 1378 AA.

ID AC Q61138;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)

DE PEGNALLY EXPRESSED PROTEIN 3 (ZN-FINGER PROTEIN PWI).

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RR STRAIN=SV129; TISSUE=LIMB BUD;

RX MEDLINE=96400442; PubMed=8806818;

RA Relaix F., Weng X., Marazzi J.G., Yang E., Copeland N., Jenkins N.,  
Spence S.E., Sassoon D.Z.  
"Pwi, a novel zinc finger gene implicated in the myogenic and neuronal  
lineages."

15  
RESULTS



[illegible]



Db 141 THWSAHGAPPTTHWSAHGAP 162

## RESULT 2

DER2\_DERPT STANDARD; PRT; 146 AA.  
 AC P49278;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE MITE ALLERGEN DER P 2 PRECURSOR (DER P II) (DPX).  
 GN DERP2.  
 OS Dermatophagoides pteronyssinus (House-dust mite).  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
 OC Acariformes; Sarcoptiformes; Astigmata; Analgoidea; Pyroglyphidae;  
 OC Dermatophagoides.  
 OX NCBI\_TaxID=6956;  
 (1)  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=90256301; PubMed=2341191;  
 RX Chua K.Y., Doyle C.R., Simpson R.J., Turner K.J., Stewart G.A.,  
 RA Thomas W.R.;  
 RT "Isolation of cDNA coding for the major mite allergen Der p II by IgE  
 RT plaque immunoassay";  
 RL Int. Arch. Allergy Appl. Immunol. 91:118-123(1990).  
 RN [2]  
 RP PARTIAL SEQUENCE OF 18-57.  
 RX MEDLINE=89278484; PubMed=2732406;  
 RA Heymann P.W., Chapman M.D., Aalberse R.C., Fox J.W.,  
 RA Platts-Mills T.A.;  
 RT "Antigenic and structural analysis of group II allergens (Der f II  
 RT and Der p II) from house dust mites (Dermatophagoides spp).";  
 RL J. Allergy Clin. Immunol. 83:1055-1067(1989).  
 RN [3]  
 RP VARIANTS.  
 RA Smith W., Hales B.J., Thomas W.R.;  
 RT "Allergens of wild house dust mites: environmental Der p 1 and Der p 2  
 RT sequence polymorphisms";  
 RL Submitted (JUN-2000) to the SWISS-PROT data bank.  
 RN [4]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE=98409423; PubMed=9737847;  
 RA Mueller G.A., Benjamin D.C., Rule G.S.;  
 RT "Tertiary structure of the major house dust mite allergen Der p 2:  
 RT sequential and structural homologies.";  
 RL Biochemistry 37:12707-12714(1998).  
 CC -!- SIMILARITY: TO MITE ALLERGEN LEP D I.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF276239; AAF86462.1; -.  
 DR PDB; 1A9V; 14-OCT-98.  
 KW Allergen; Signal; 3D-structure; Polymorphism.  
 FT SIGNAL 1 17 MITE ALLERGEN DER P 2.  
 FT CHAIN 18 146  
 FT DISULFID 25 136  
 FT DISULFID 38 44  
 FT DISULFID 90 95  
 FT VARIANT 39 39  
 FT VARIANT 40 40  
 FT VARIANT 44 44  
 FT VARIANT 47 47  
 FT VARIANT 49 49  
 FT VARIANT 56 56  
 FT VARIANT 61 61  
 FT VARIANT 75 75  
 FT VARIANT 78 78  
 FT H -> A.  
 FT G -> L.  
 FT C -> N.  
 FT H -> S.  
 FT A -> T.  
 FT A -> Y.  
 FT N -> L.  
 FT I -> Y.  
 FT L -> C.

FT VARIANT 81 81 D -> V.  
 FT VARIANT 95 95 C -> P.  
 FT VARIANT 98 98 V -> T.  
 FT VARIANT 108 108 T -> V.  
 FT VARIANT 111 111 V -> L.  
 FT VARIANT 114 114 I -> N.  
 FT VARIANT 116 116 P -> A.  
 FT VARIANT 118 118 S -> A.  
 FT VARIANT 127 127 V -> L.  
 SQ SEQUENCE 146 AA; 15999 MW; 591B2FA7FD26D3AF CRC64;

Query Match 12.5%; Score 99; DB 1; Length 146;  
 Best Local Similarity 100.0%; Pred.No. 0.0019;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 GCHGSEPCIHRGKPF 67

Db 37 GCHGSEPCIHRGKPF 52

## RESULT 3

DEF2\_DERFA STANDARD; PRT; 146 AA.  
 ID DEF2\_DERFA Q00855; P39672; Q26359;  
 AC Q00855; P39672; Q26359;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE MITE ALLERGEN DER F 2 PRECURSOR (DER F II).  
 GN DERF2.  
 OS Dermatophagoides farinae (House-dust mite).  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
 OC Acariformes; Sarcoptiformes; Astigmata; Analgoidea; Pyroglyphidae;  
 OC Dermatophagoides.  
 OX NCBI\_TaxID=6954;  
 (1)  
 RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RP MEDLINE=91291341; PubMed=1368682;  
 RX Yuuki T., Okumura Y., Ando T., Yamakawa H., Suko M., Haida M.,  
 RA Okudaira H.;  
 RT "Cloning and expression of cDNA coding for the major house dust mite  
 RT allergen Der f II in Escherichia coli.";  
 RL Agric. Biol. Chem. 55:1233-1238(1991).  
 RN [2]  
 RP SEQUENCE OF 4-146 FROM N.A.  
 RX MEDLINE=94256850; PubMed=8198452;  
 RA Okuhira H.;  
 RT "Molecular biology of mite antigens";  
 RL Arerugi 43:435-440(1994).  
 RN [3]  
 RN DISULFIDE BONDS, AND PARTIAL SEQUENCE.  
 RX MEDLINE=93283958; PubMed=8508052;  
 RA Nishiyama C., Yuuki T., Takai T., Okumura Y., Okudaira H.;  
 RT "Determination of three disulfide bonds in a major house dust mite  
 RT allergen, Der f II.";  
 RL Int. Arch. Allergy Immunol. 101:159-166(1993).  
 RN [4]  
 RN PARTIAL SEQUENCE OF 18-52.  
 RP MEDLINE=89278484; PubMed=2732406;  
 RX Heymann P.W., Chapman M.D., Aalberse R.C., Fox J.W.,  
 RA Platts-Mills T.A.;  
 RT "Antigenic and structural analysis of group II allergens (Der f II  
 RT and Der p II) from house dust mites (Dermatophagoides spp).";  
 RL J. Allergy Clin. Immunol. 83:1055-1067(1989).  
 RN [5]  
 RN STRUCTURE BY NMR.  
 RP MEDLINE=98079088; PubMed=9417088;  
 RA Ichikawa S., Hatanaka H., Yuuki T., Iwamoto N., Kojima S.,  
 RA Nishiyama C., Ogura K., Okumura Y., Inagaki F.;  
 RT "Solution structure of Der f 2, the major mite allergen for atopic  
 RT diseases.";  
 RL J. Biol. Chem. 273:356-360(1998).  
 CC -!- MISCELLANEOUS: THE SEQUENCE SHOWN HERE IS FROM CLONE 2. THE

CC N-TERMINAL SEQUENCE (AA 1-8) FROM CLONE 1 AND 11 ARE NOT YET  
 CC KNOWN.  
 CC -1- SIMILARITY: TO MITE ALLERGEN LEP D I.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D10447; BAA01239.1; -  
 DR EMBL; D10448; BAA01240.1; -  
 DR EMBL; D10449; BAA01241.1; -  
 DR EMBL; S70378; AAB30829.1; -  
 DR PIR; PS0417; PS0417.  
 DR PDB; 1AHK; 08-APR-98.  
 DR PDB; 1AHM; 08-APR-98.  
 KW Allergen; Signal; Polymorphism; 3D-structure.  
 FT SIGNAL 1 17  
 FT CHAIN 18 146 MITE ALLERGEN DER F 2.  
 FT DISULFID 25 136  
 FT DISULFID 38 44  
 FT DISULFID 90 95  
 FT VARIANT 93 93 M -> V (IN CLONE 1).  
 FT VARIANT 105 105 I -> A (IN CLONE 11).  
 FT VARIANT 128 128 I -> V (IN CLONE 11).  
 FT VARIANT 142 142 G -> A (IN CLONE 11).  
 FT CONFLICT 5 8 ILCL -> GTMV (IN REF. 2).  
 FT SEQUENCE 146 AA; 15802 MW; FAL18206CD88534A CRC64;  
 CC  
 CC Query Match 12.28; Score 97; DB 1; Length 146;  
 CC Best Local Similarity 88.28; Pred. No. 0.003; Indels 0; Gaps 0;  
 CC Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 QY 52 GCHGSEPCIIHRGKPFSS 68  
 DB 37 GCHGSDPCIIHRGKPF 53  
 CC  
 RESULT 4  
 BAR3\_CHITE  
 ID BAR3\_CHITE STANDARD; PRT; 1700 AA.  
 AC Q03376;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-OCT-1994 (Rel. 30, Last annotation update)  
 DE BALBIANI RING PROTEIN 3 PRECURSOR.  
 GN BR3.  
 OS Chironomus tentans (Midge).  
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;  
 CC Chironomidae; Chironominae; Chironomus.  
 CC NCBI\_TaxID=7153;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=90172404; PubMed=1689777;  
 RA Paulsson G., Lendahl U., Galli J., Ericsson C., Wileslander L.;  
 RT "The Balbiani ring 3 gene in Chironomus tentans has a diverged  
 RT repetitive structure split by many introns.";  
 RL J. Mol. Biol. 211:331-349(1990).  
 CC -1- FUNCTION: USED BY THE LARVAE TO CONSTRUCT A SUPRAMOLECULAR  
 CC STRUCTURE, THE LARVAL TUBE. BALBIANI RING PROTEIN 3 COULD PLAY A  
 CC ROLE AS A TRANSPORT PROTEIN THAT BINDS TO OTHER PROTEINS  
 CC INTRACELLULARLY AND IN THE GLAND LUMEN IN ORDER TO PREVENT THESE  
 CC FROM FORMING WATER-INSOLUBLE FIBERS TOO EARLY.  
 CC -1- SUBCELLULAR LOCALIZATION: SECRETED.  
 CC -1- TISSUE SPECIFICITY: SALIVARY GLAND.  
 CC -1- DOMAIN: HAS 82 APPROXIMATE REPEATS OF CYS-X-CYS-X-CYS.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X52263; CAA36506.1; -  
 DR PIR; S08167; S08167.  
 DR HSP; P18055; 2MRB.  
 DR InterPro; IPR000853; -  
 DR PRINTS; PR00876; MTNEMATODE.  
 KW Repeat; Signal.  
 FT SIGNAL 1 ? POTENTIAL.  
 FT CHAIN ? 1700 BALBIANI RING PROTEIN 3.  
 FT SEQUENCE 1700 AA; 186145 MW; 34202B28521B0815 CRC64;  
 CC  
 CC Query Match 11.38; Score 90; DB 1; Length 1700;  
 CC Best Local Similarity 22.88; Pred. No. 0.16;  
 CC Matches 37; Conservative 10; Mismatches 49; Indels 66; Gaps 7;  
 CC  
 QY 35 SCHGSEPCIIHRGKPFSSCHGSE-----PCIIHRGKPFSSCHGSE----- 74  
 DB 1152 SKCKNP-----KPANGCTGVQEWNEEKQCCEPKDKPKKQCGQDWNHHCQCGCPT 1205  
 QY 75 -----PCIIHRGKPFSSCHGSE-----PCIIHRGKPFSSCHGSE----- 105  
 DB 1206 PAPTCSNNOKYSNVSCGCGNPKGKPGNQIWCNTCRVCPCPKMEXPADNCKTKW 1265  
 QY 106 GSE-PCIIHRGKPFSSCHG-----SEPCIIHRGKPFSS 136  
 DB 1266 NDEMCCVCKPCGCEGCKGVMKNANTSCCECPADKAKPAS 1307  
 CC  
 RESULT 5  
 Z071\_XENLA  
 ID Z071\_XENLA STANDARD; PRT; 898 AA.  
 AC P18751;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-OCT-1994 (Rel. 30, Last annotation update)  
 DE OOCYTE ZINC FINGER PROTEIN XLCOF7.1 (FRAGMENT).  
 OS Xenopus laevis (African clawed frog).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
 CC Xenopodinae; Xenopus.  
 CC NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE OF 1-311 FROM N.A.  
 RX MEDLINE=89345612; PubMed=2503827;  
 RA Knoechel W., Poeting A., Koester M., el Baradi T., Niefeld W.,  
 RA Bouwmeester T., Pieler T.;  
 RT "Evolutionary conserved modules associated with zinc fingers in  
 RT Xenopus laevis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:6097-6100(1989).  
 RN [2]  
 RP SEQUENCE OF 284-898 FROM N.A.  
 RX MEDLINE=90040698; PubMed=2509712;  
 RA Niefeld W., el-Baradi T., Mentzel H., Pieler T., Koester M.,  
 RA Poeting A., Knoechel W.;  
 RT "Second-order repeats in Xenopus laevis finger proteins.";  
 RL J. Mol. Biol. 208:639-659(1989).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

```

DR EMBL; M25866; AAA50013.1; -.
DR PIR; A33282; A33282.
DR PIR; S06546; S06546.
DR HSSP; P08047; 1SP2.
DR InterPro; IPR000822; -.
DR Pfam; PF00096; zf-C2H2; 21.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 21.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 21.
KW zinc-finger; Metal-binding; DNA-binding; Repeat.
FT ZN_FING 289 311
FT ZN_FING 342 364
FT ZN_FING 370 392
FT ZN_FING 398 420
FT ZN_FING 425 448
FT ZN_FING 426 448
FT ZN_FING 454 476
FT ZN_FING 481 504
FT ZN_FING 482 504
FT ZN_FING 510 532
FT ZN_FING 537 561
FT ZN_FING 568 589
FT ZN_FING 567 589
FT ZN_FING 595 617
FT ZN_FING 623 645
FT ZN_FING 651 673
FT ZN_FING 679 701
FT ZN_FING 707 729
FT ZN_FING 735 758
FT ZN_FING 764 786
FT ZN_FING 792 814
FT ZN_FING 820 842
FT ZN_FING 848 870
FT ZN_FING 876 898
FT ZN_FING 898 898
FT NON_TER 898 898
SQ SEQUENCE 898 AA; 101167 MW; 795D806E569E80B6 CRC64;

Query Match 11.1%; Score 88; DB 1; Length 898;
Best Local Similarity 24.3%; Pred. No. 0.13;
Matches 44; Conservative 13; Mismatches 60; Indels 64; Gaps 12;

QY 19 QYIKANSKFGITELSSCHGSEPCI-----IHRG-KPF-----GGCHGSEPCIHR----- 63
DB 614 QMTHGKPFSCSECKGCFASSSDLTFFHRTHTGKPFSCSECKGKYSKSLVHHQRT 673

QY 64 --GKPFSSCHGSEPCI-----IHRGKPFGGCHGSEPC-----IHRGK 99
DB 674 TGEKPF-SCSKCDKCPASSSELNIHQRTHTGKAFSCSECKGCTTNRSQLSRHQMTHTGE 732

QY 100 PFSSCHGSEPCI-----HR-GKPF-----GGCHGSEPC-----CIHRG-KPF 135
DB 733 KPISCECECFVSSQLTAHQQAHRMVKPFSCLECKGCFNSRNFARHQMTHTGKPF 792

QY 136 S 136
DB 793 S 793

RESULT 6
ZF92_MOUSE
ID ZF92_MOUSE STANDARD; PRT; 488 AA.
AC Q62396;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ZINC FINGER PROTEIN 92 (ZFP-92).
GN ZFP92.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.

TISSUE=Embryo;
MEDLINE=96425694; PubMed=8828036;
RA Levin M.L., Chatterjee A., Pragila A., Worley K.C., Wehnert M.,
RA Zhuchenko O., Smith R.F., Lee C.C., Herman G.E.;
RT "A comparative transcription map of the murine bare patches (Bpa) and
RT striated (Str) critical regions and human Xq28.";
RL Genome Res. 6:465-477(1996).
CC -!- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -!- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; U47104; AAC52629.1; -.
CC MGD; MGI:108094; Zfp92.
CC InterPro; IPR000822; -.
CC InterPro; IPR001909; -.
CC Pfam; PF01352; KRAB; 1.
CC Pfam; PF00096; zf-C2H2; 9.
CC PRINTS; PR00048; ZINC_FINGER.
CC PROSITE; PS00805; KRAB; 1.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 9.
CC Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 14 85
FT DOMAIN 141 432
FT ZN_FING 141 163
FT ZN_FING 169 191
FT ZN_FING 197 219
FT ZN_FING 225 247
FT ZN_FING 253 275
FT ZN_FING 281 303
FT ZN_FING 337 359
FT ZN_FING 410 432
FT ZN_FING 488 AA; 55961 MW; 6A649E30F2043699 CRC64;

Query Match 10.8%; Score 86; DB 1; Length 488;
Best Local Similarity 33.6%; Pred. No. 0.12;
Matches 49; Conservative 10; Mismatches 39; Indels 48; Gaps 15;

QY 14 TELGGQYIKANSKFGITELSSCH-GSEPCIHR-GKPFGG---CHGSEPCIHRG-KP 66
DB 256 TECG---KAFSRSNLIEHQRIHSGQPKYICKGKAFKGVSVQVHHQ---LIHRGDKP 308

QY 67 FSSCHGSEPCIHRGKPFGGCHG-SEPCIHRG-KPF---SSCHGSEPCIHRGKPFGG--- 120
DB 309 F-TCH-----EYGAFAFGSLGLSQHQRVHRGKPFYECSEC-----GRAFGRA 350

QY 121 -----CHGSEPCIHR--RGKPFGR 137
DB 351 NLFKHQVHVGVR-LQHRTRGKGFQR 375

RESULT 7
SP87_DICDI
ID SP87_DICDI STANDARD; PRT; 677 AA.
AC P54643;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SP87 COAT PROTEIN SP87 PRECURSOR (PL3 PROTEIN).
GN PSPD.
OS Dictyostelium discoideum (slime mold).

```





```

Query Match      10.2%; Score 80.5; DB 1; Length 197;
Best Local Similarity 27.1%; Pred. No. 0.17;
Matches 51; Conservative 11; Mismatches 51; Indels 75; Gaps 15;

QY 9 KFIGITELGGQYIKAN-----SKFIGITELSSCHGSEPCI-----IHRG-KPFG 51
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 4 KSISSCGCGCFIKSELTVHOMTHIGEKTYSCSECGKCFASLSLHRVHQIHTGEKPF- 62
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 52 GCHGSEPCIHRG-----KPF--SSCHGSEPCIHRGKPGCGCHGSEPCIHR 97
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 63 SCSCGCGKFLNRLVHRHRTHTGKPFPCSEC-----GRFA---ASSDLRVRH 109
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 98 -----GRPFSSCHGSEPCI-----HRG-KPF-----GCHG-SEPCIHR--- 131
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 110 RTHTGERPF-SCSECKERFLNPSLVRHYRTHGKPFSCSECGKCFARSSDLTVHRRS 168
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 132 ---CKPFS 136
| | | | |
DB 169 HTREKPF 176
| | | | |

RESULT 11
M84A_DROME STANDARD; PRT; 63 AA.
AC Q01642; Q9VIA3;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MALE SPECIFIC SPERM PROTEIN MST84DA.
GN MST84DA OR CG17946.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Sutton G.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA George R.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.F., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fouts D., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

```

```

RA Merkulov G., Milshina N.V., Mobarriy C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasmann D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -!- DEVELOPMENTAL STAGE: PRIMARY SPERMATOCYTES.
CC -!- DOMAIN: THIS PROTEIN IS MOSTLY COMPOSED OF REPETITIVE C-G-P
CC MOTIFS.
CC -!- SIMILARITY: BELONGS TO THE MST(3)CGP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X67703; CAA47937.1; -.
DR EMBL; AE003672; AAF54022.1; -.
DR HSSP; P01180; INPO.
DR FlyBase; FBgn004172; Mst84Da.
KW Spermatogenesis; Repeat; Multigene family.
SQ SEQUENCE 63 AA; 5806 MW; BF84CD74CBCEFD3F CRC64;

Query Match      10.0%; Score 79; DB 1; Length 63;
Best Local Similarity 29.7%; Pred. No. 0.083;
Matches 27; Conservative 2; Mismatches 30; Indels 32; Gaps 6;

QY 20 YTKANSKFTGITELSSCHGSEPCIHRGKPGCGCHGSEPCIHRGKPFSSCHGSEPCIH 79
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 4 YVNPYVLVGGPCGCGGCGPC-----GGC---GPC-----CGGCGGCC-- 40
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 80 RCKPFGCGCHGSEPCIHRGKPFSSCHGSEPC 110
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 41 -----GPCGCGGCC---GGTSSFC-GCGPC 62
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 12
CSP_PLAKU STANDARD; PRT; 351 AA.
AC P04922;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).
OS Plasmodium knowlesi (strain nuri).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5852;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85272582; PubMed=4023712;
RA Sharma S., Svec P., Mitchell G.H., Godson G.N.;
RT "Diversity of circumsporozoite antigen genes from two strains of the
RT malarial parasite Plasmodium knowlesi."
RL Science 229:779-782(1985).
CC -!- MISCELLANEOUS: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC SURFACE ANTIGEN ON THE SPOOROZOITE (THE INFECTIVE STAGE OF THE
CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC VERTEBRATE HOST).
CC -!- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR

```

CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES  
 CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.  
 CC -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M11031; AAA29540.1; -  
 CC PIR: A26253; OZZORU.  
 CC InterPro: IPR000884; -  
 CC Pfam: PF00090; tsp\_1; 1.  
 CC PRINTS: PR01303; CRMSPRZOITE.  
 KW Malaria; Sporozoite; Repeat; Signal.  
 FT SIGNAL 1 19 PROBABLE.  
 FT CHAIN 20 351 CIRCUMSPOROZOITE PROTEIN.  
 FT DOMAIN 101 235 15 X 9 AA TANDEM REPEATS OF A-A-G-A-G-G-  
 FT E-Q-P.  
 SQ SEQUENCE 351 AA; 34782 MW; A85E87A152E6485B CRC64;  
 Query Match 10.0%; Score 79; DB 1; Length 351;  
 Best Local Similarity 27.2%; Pred. No. 0.43;  
 Matches 28; Conservative 13; Mismatches 48; Indels 14; Gaps 5;  
 QY 38 GSEPCIHRG-KPFGCHGSEPCIHRG-KPSSCHGSEPCIHRG-KPFGCHGSEPCI 94  
 DB 124 GEQPAAGARGEQPAAGAGGQPAAGAGGQPAAGAGGQPAAGAGGQPAAGAGGQPA 183  
 QY 95 IHRG-----KPSSCHGSEPCIHRG-KPFGCHGSEPCI 126  
 DB 184 GARGEQPAAGAGGQPAAGAGGQPAAGAGGQPAAGAGGQPAAGAGGQPAAGAGGQ 226  
 RESULT 13  
 ITB8 HUMAN  
 ID ITB8 HUMAN STANDARD; PRT; 769 AA.  
 AC P26012;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE INTEGRIN BETA-8 PRECURSOR.  
 GN ITGB8.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=92011767; PubMed=1918072;  
 RA Moyle M., Napier M.A., McLean J.W.;  
 RT "Cloning and expression of a divergent integrin subunit beta 8.";  
 RL J. Biol. Chem. 266:19650-19658(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Tin-william A., Sutterer C., Fronick B.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: INTEGRINS ARE A FAMILY OF TRANSMEMBRANE COMPLEXES THAT  
 CC FUNCTION AS RECEPTORS FOR CELL ADHESION MOLECULES.  
 CC -!- SUBUNIT: DIMER OF AN ALPHA AND BETA SUBUNIT. BETA-8 ASSOCIATES  
 CC WITH ALPHA-V.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -!- TISSUE SPECIFICITY: PLACENTA, KIDNEY, BRAIN, OVARY, UTERUS, AND IN  
 CC SEVERAL TRANSFORMED CELLS. TRANSIENTLY EXPRESSED IN 293 HUMAN  
 CC EMBRYONIC KIDNEY CELLS.  
 CC -!- PTM: THE CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN DISULFIDE  
 CC BONDS.

CC -!- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M73780; AAA36034.1; -  
 CC PIR: AC004130; AAC01769.1; -  
 CC MIM: 604160;  
 CC InterPro: IPR000561; -  
 CC InterPro: IPR001169; -  
 CC InterPro: IPR002369; -  
 CC Pfam: PF00362; Integrin\_B; 1.  
 CC PROSITE: PS00243; INTEGRIN\_BETA; 2.  
 CC PROSITE: PS0022; EGF\_1; UNKNOWN\_1.  
 CC PROSITE: PS01186; EGF\_2; UNKNOWN\_1.  
 KW Integrin; Cell adhesion; Transmembrane; Glycoprotein; Repeat;  
 KW Extracellular matrix; Cytoskeleton; Signal.  
 FT SIGNAL 1 42 POTENTIAL.  
 FT CHAIN 43 769 INTEGRIN BETA-8.  
 FT DOMAIN 43 684 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 685 704 POTENTIAL.  
 FT DOMAIN 705 769 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 471 629 CYSTEINE-RICH REPEATS.  
 FT REPEAT 471 510 I.  
 FT REPEAT 511 552 II.  
 FT REPEAT 553 592 III.  
 FT REPEAT 593 629 IV.  
 FT CARBOHYD 233 233 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 402 402 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 421 421 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 431 431 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 456 456 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 466 466 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 648 648 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 769 AA; 85631 MW; F7E399A92B12A65 CRC64;  
 Query Match 10.0%; Score 79; DB 1; Length 769;  
 Best Local Similarity 25.8%; Pred. No. 0.9;  
 Matches 42; Conservative 14; Mismatches 55; Indels 52; Gaps 11;  
 QY 9 KFIGITELGQYIKAN-----SKFIGITELSS-----CH-----GSEP 41  
 DB 451 KPIGFNETAKIHRNCSQCEDNRGPKGVDETFDLSKFCQCDENKCHDFDQFSSES 510  
 QY 42 CIHRGKPGGCHGSEPCIHRGKPFSSCHGSEPCIHRGKPFSG-CHGSE-PCIHRGK 99  
 DB 511 CKSHKQDPV--CSGRGVQVC--GR--CSCHK-----TKLGVYKGYCKEKKDFSCPYHHG- 558  
 QY 100 PFSSCHGSEPCIHRGKPFGGCHG-----SEPCIHRGK 133  
 DB 559 --NLCAGHGCEAGRCQCFSSWEGDRCCQPSAAAHCVNSKGQ 599  
 RESULT 14  
 TETX\_CLOTE  
 ID TETX\_CLOTE STANDARD; PRT; 1314 AA.  
 AC P04958;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE TETANUS TOXIN PRECURSOR (EC 3.4.24.68) (TENTOXILYSIN).  
 OS Clostridium tetani.  
 OG Plasmid.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1513;

```

CC CC  GANGLIOSIDE RECEPTORS.
CC CC  -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
CC CC  METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
CC CC  SUBFAMILY.
CC CC  -----
CC CC  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC  the European Bioinformatics Institute. There are no restrictions on its
CC CC  use by non-profit institutions as long as its content is in no way
CC CC  modified and this statement is not removed. Usage by and for commercial
CC CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC  or send an email to license@isb-sib.ch).
CC CC  -----
CC CC  EMBL; X04436; CAA28033.1; -
CC CC  EMBL; M12739; AAA23282.1; -
CC CC  EMBL; X06214; CAA29564.1; -
CC CC  PIR; A25689; BTCLTN.
CC CC  PDB; 1AF9; 29-APR-98.
CC CC  PDB; 1A8D; 14-OCT-98.
CC CC  MEROPS; M27.001; -.
CC CC  InterPro; IPR000130; -.
CC CC  InterPro; IPR000395; -.
CC CC  Pfam; PF01742; Peptidase_M27; 1.
CC CC  PRINTS; PR00760; BONTOKILYSIN
CC CC  PROSITE; PS00142; ZINC_PROTEASE; 1.
CC CC  Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
CC CC  3D-structure.
CC CC  INIT_MET 0 0
CC CC  FT CHAIN 1 456 TETANUS TOXIN LIGHT CHAIN.
CC CC  FT CHAIN 457 1314 TETANUS TOXIN HEAVY CHAIN.
CC CC  FT METAL 232 233 ZINC (CATALYTIC) (BY SIMILARITY).
CC CC  FT ACT_SITE 233 233 BY SIMILARITY.
CC CC  FT METAL 236 236 ZINC (CATALYTIC) (BY SIMILARITY).
CC CC  FT TRANSMEM 226 246 POTENTIAL.
CC CC  FT TRANSMEM 669 689 POTENTIAL.
CC CC  FT DISULFID 438 466 INTERCHAIN.
CC CC  FT DISULFID 1076 1092
CC CC  SQ SEQUENCE 1314 AA; 150550 MW; 134C3657133EF81D CRC64;

Query Match 9.8%; Score 78; DB 1; Length 1314;
Best Local Similarity 69.2%; Pred. No. 1.9;
Matches 18; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 2 QYIKANSFGICITELGGQYIKANSKF 27
| | | | | | | | | | | | | | | |
Db 829 QYIKANSFGICITELKKLESINKVF 854

RESULT 15
INSR_DROME
ID INSR_DROME STANDARD; PRT; 2146 AA.
AC P09208; Q24089; Q24023;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE INSULIN-LIKE RECEPTOR PRECURSOR (EC 2.7.1.112).
GN INSR OR INR-A OR DIR-A.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95354655; PubMed=7628438;
RA Fernandez R., Tabarini D., Azpiazu N., Frasch M., Schlessinger J.;
RT "The Drosophila insulin receptor homolog: a gene essential for
RT embryonic development encodes two receptor isoforms with different
RT signaling potential.";
RL EMEQ J. 14:3373-3384(1995).
RN [2]
RP SEQUENCE FROM N.A.

```





QY 111'-----IIHRKPFQGC---HGSEPCII 129  
Db 646 ECSSGLIDSLERAREFRGCTIITGTEPLTI 675

Search completed: April 14, 2001, 10:17:43  
Job time: 345 sec

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	101	12.7	1571	2	T14155	zinc finger protein	
2	99	12.5	146	2	A60381	allergen Der p II	
3	97	12.2	129	2	A61501	allergen Der f II	
4	97	12.2	129	2	JU0394	allergen Der f II	
5	97	12.2	138	2	B61241	allergen Der f II	
6	97	12.2	138	2	A61241	allergen Der f II	
7	90	11.3	1700	2	S08167	Balbani ring 3 pr	
8	88	11.1	615	2	S08546	finger protein (cl	
9	85	10.7	169	1	S18946	ultra high-sulfur	
10	85	10.7	232	1	T31524	hypothetical prote	
11	84	10.6	230	2	A38346	ultra-high-sulfur	
12	82.5	10.4	1378	2	T30173	zinc finger protein	
13	82	10.3	102	2	A72507	hypothetical prote	
14	81.5	10.3	1584	2	T26674	hypothetical prote	
15	81	10.2	846	2	H70599	hypothetical prote	
16	80.5	10.2	197	2	S06564	finger protein (cl	
17	80.5	10.2	1797	2	A55677	laminin beta-2 cha	
18	80	10.1	361	2	T14460	oleosin homolog gb	
19	80	10.1	1523	2	T13953	MEGF5 protein - ra	
20	79	10.0	63	2	S25772	testis-specific pr	
21	79	10.0	351	1	OZZQKU	circumsporozoite p	
22	79	10.0	769	2	A41029	integrin beta-8 ch	
23	78	9.8	1315	1	BTCLTN	tentoxylysin (EC 3	
24	78	9.8	1820	2	A55494	latent transformin	
25	78	9.8	2101	2	S57245	insulin receptor (	
26	78	9.8	2148	1	A56081	insulin receptor -	
27	77.5	9.8	471	2	T33997	hypothetical prote	
28	77.5	9.8	1801	1	MMRTS	laminin beta-2 cha	
29	76.5	9.6	3871	2	T22812	hypothetical prote	

F;1-17/Domain: signal sequence #status predicted <SIG>  
F;18-146/Product: allergen Der p II #status predicted

Query Match 12.5%; Score 99; DB 2; Length 146;  
Best Local Similarity 100.0%; Pred. No. 0.0053;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 GCHGSEPCIIHRGKPF 67  
Db 37 GCHGSEPCIIHRGKPF 52

RESULT 3  
A:Residues: 1-129 <YU>  
A:Superfamily: allergen Der p II  
C:Species: Dermatophagoides farinae (fragment)  
C:Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 13-Sep-1998  
C:Accession: A61501  
R:Trudinger, M.; Chua, K.Y.; Thomas, W.R.  
Clin. Exp. Allergy 21, 33-37, 1991  
A:Title: cDNA encoding the major mite allergen Der f II.  
A:Reference number: A61501; MUID:91215495  
A:Accession: A61501  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-129 <TR>  
C:Superfamily: allergen Der p II

Query Match 12.2%; Score 97; DB 2; Length 129;  
Best Local Similarity 88.2%; Pred. No. 0.0074;  
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 52 GCHGSEPCIIHRGKPF 68  
Db 20 GCHGSDPCIIHRGKPF 36

RESULT 4  
A:Residues: 1-138 <YU>  
A:Superfamily: allergen Der p II  
C:Species: Dermatophagoides farinae (fragment)  
C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 17-Mar-1999  
C:Accession: J00394  
R:Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haida, M.; Okudaira, H.  
Agric. Biol. Chem. 55, 1233-1238, 1991  
A:Title: Cloning and expression of cDNA coding for the major house dust mite allergen Der f II.  
A:Reference number: PS0417; MUID:91291341  
A:Accession: J00394  
A:Molecule type: mRNA  
A:Residues: 1-129 <YU>  
C:Superfamily: allergen Der p II

Query Match 12.2%; Score 97; DB 2; Length 129;  
Best Local Similarity 88.2%; Pred. No. 0.0074;  
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 52 GCHGSEPCIIHRGKPF 68  
Db 20 GCHGSDPCIIHRGKPF 36

RESULT 5  
A:Residues: 1-138 <YU>  
A:Superfamily: allergen Der p II  
C:Species: Dermatophagoides farinae (fragment)  
C:Date: 12-May-1994 #sequence\_revision 27-Jun-1994 #text\_change 13-Sep-1998  
C:Accession: B61241; J00395  
R:Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haida, M.; Dohi, M.; Okudaira, H.  
Int. Arch. Allergy Appl. Immunol. 94, 354-356, 1991  
A:Title: Synthesis of biologically active recombinant Der f II.  
A:Reference number: A61241; MUID:92040281  
A:Accession: B61241  
A:Molecule type: mRNA

A:Residues: 1-138 <YU>  
C:Superfamily: allergen Der p II  
F:1-9/Domain: signal sequence (fragment) #status predicted <SIG>  
F:10-138/Product: allergen Der f II #status predicted <MAT>

Query Match 12.2%; Score 97; DB 2; Length 138;  
Best Local Similarity 88.2%; Pred. No. 0.0079;  
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 52 GCHGSEPCIIHRGKPF 68  
Db 29 GCHGSDPCIIHRGKPF 45

RESULT 6  
A:Residues: 1-138 <YU>  
A:Superfamily: allergen Der f II precursor - house-dust mite (Dermatophagoides farinae) (fragment)  
C:Species: Dermatophagoides farinae  
C:Date: 12-May-1994 #sequence\_revision 27-Jun-1994 #text\_change 13-Sep-1998  
C:Accession: A61241; PS0417  
R:Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haida, M.; Dohi, M.; Okudaira, H.  
Int. Arch. Allergy Appl. Immunol. 94, 354-356, 1991  
A:Title: Synthesis of biologically active recombinant Der f II.  
A:Reference number: A61241; MUID:92040281  
A:Accession: A61241  
A:Molecule type: mRNA  
A:Residues: 1-138 <YU>  
A:Note: part of this sequence, including the amino end of the mature protein, was conserved in the mature protein.  
C:Superfamily: allergen Der p II  
F:1-9/Domain: signal sequence (fragment) #status predicted <SIG>  
F:10-138/Product: allergen Der f II #status experimental <MAT>

Query Match 12.2%; Score 97; DB 2; Length 138;  
Best Local Similarity 88.2%; Pred. No. 0.0079;  
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 52 GCHGSEPCIIHRGKPF 68  
Db 29 GCHGSDPCIIHRGKPF 45

RESULT 7  
A:Residues: 1-138 <YU>  
A:Superfamily: allergen Der p II  
C:Species: Chironomus tentans  
C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 21-Jul-2000  
C:Accession: S08167  
R:Paulsson, G.; Lendahl, U.; Gallii, J.; Ericsson, C.; Wieslander, L.  
J. Mol. Biol. 211, 331-349, 1990  
A:Title: The balbiani ring 3 gene in Chironomus tentans has a diverged repetitive structure.  
A:Reference number: S08167; MUID:90172404  
A:Accession: S08167  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-1700 <PAU>  
A:Cross-references: GB:X52263; NID:g7057; PIDN:CAA36506.1; PID:g7058  
C:Genetics: BR3  
A:Gene: BR3  
A:Map position: 4  
C:Superfamily: unassigned Balbiani ring proteins

Query Match 11.3%; Score 90; DB 2; Length 1700;  
Best Local Similarity 22.8%; Pred. No. 0.43;  
Matches 37; Conservative 10; Mismatches 49; Indels 66; Gaps 7;

QY 35 SCHGSEPCIIHRGKPF 74  
Db 1152 SCCKNP-----KPANGCTGVQEWNEEKQCCEPKDKPKKQCPGQDWNHHCQCQCPT 1205

QY 75 -----PCIIHRGKPF 105

```
Db 1206 PAPTCSNNQKYSNVSCGNCNPKXNGCNGNIWCDNTCRCVCPKMKRXPADNCKTKWW 1265
QY 106 GSE-PCIIHRGKPFPGCHG-----SEPCIIHRGKPF 136
Db 1266 NDEMCCVCVKPGCGEGCKGVMKWNANTSCCEPADKAKPAS 1307

RESULT 8
S06546
finger protein (clone Xlcof7.1) - African clawed frog (fragment)
C:Species: Xenopus laevis (African clawed frog)
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 01-Dec-2000
C:Accession: S06546
R:Nietfeld, W.; El-Baradi, T.; Mentzel, H.; Pieler, T.; Koester, M.; Poeting, A.; Knoech
J. Mol. Biol. 208, 639-659, 1989
A:Title: Second-order repeats in Xenopus laevis finger proteins.
A:Reference number: S05632; MUID:90040698
A:Accession: S06546
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-615 <NIE>
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; zinc finger

Query Match 11.1%; Score 88; DB 2; Length 615;
Best Local Similarity 24.3%; Pred. No. 0.25;
Matches 44; Conservative 13; Mismatches 60; Indels 64; Gaps 12;

QY 19 QYTKANSKFGITELSSCHGSEPCI-----IHRG-KPF-----GGCHGSEPCIHR----- 63
Db 331 QMPHTGKPFSCGCKGCFASSDLTFPHRTHTGKPFSCGCKGCKYKSSSLVHHQRT 390

QY 64 --GKPFSSCHGSEPCI-----IHRGKPFPGCHGSEPC-----IHRGK 99
Db 391 TGSKPF-SCSKCDKCFASSSELNIHQTHTGKAFSCGCKPNTNRSQLSRHOIMHTGE 449

QY 100 PFSSCHGSEPCI-----HR-GKPF-----GGCHGSEPC-----CIIHRG-KPF 135
Db 450 KPISCECECFVSSQLTAHQOAHMWVAFSCLECGCKCFNSRNPFAHQMIHTGKPF 509

QY 136 S 136
Db 510 S 510

RESULT 9
S18946
ultra high-sulfur keratin 1 - human
N:Alternate names: UHS keratin; ultra high-sulfur matrix protein
C:Species: Homo sapiens (man)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: S18946; B36686
R:Drabant, B.; Doenecke, D.
submitted to the EMBL Data Library, December 1991
A:Description: Nucleotide sequence of a Human high-sulphur keratin cDNA.
A:Reference number: S18946
A:Accession: S18946
A:Molecule type: mRNA
A:Residues: 1-169 <DRA>
A:Cross-references: EMBL:X63755; NID:g32471; PIDN:CAA45283.1; PID:g32472
R:MacKinnon, P.J.; Powell, B.C.; Rogers, G.E.
J. Cell Biol. 111, 2587-2600, 1990
A:Title: Structure and expression of genes for a class of cysteine-rich proteins of the
A:Reference number: A36686; MUID:91115951
A:Accession: B36686
A:Molecule type: DNA
A:Residues: 1-39, 'Y', 41-169 <MAC>
A:Cross-references: GB:X55293; NID:g34078; PIDN:CAA39005.1; PID:g34079
C:Genetics:
A:Gene: GDB:KRNL
A:Cross-references: GDB:125257; OMIM:148021
```

```
A:Map position: 11q13-11q13
C:Superfamily: ultra-high-sulfur keratin
C:Keywords: hair; tandem repeat
F:7-15/Region: Ser-rich nonapeptide repeat
F:59-68/Region: Gly-rich decapeptide repeat
F:69-78/Region: Gly-rich decapeptide repeat
F:79-88/Region: Cys-rich decapeptide repeat
F:89-97/Region: Ser-rich nonapeptide repeat
F:98-107/Region: Cys-rich decapeptide repeat
F:108-117/Region: Cys-rich decapeptide repeat
F:118-126/Region: Ser-rich nonapeptide repeat
F:127-136/Region: Cys-rich decapeptide repeat
F:137-145/Region: Ser-rich nonapeptide repeat
F:146-155/Region: Cys-rich decapeptide repeat
F:156-165/Region: Cys-rich decapeptide repeat

Query Match 10.7%; Score 85; DB 1; Length 169;
Best Local Similarity 27.9%; Pred. No. 0.14;
Matches 34; Conservative 9; Mismatches 53; Indels 26; Gaps 8;

QY 12 GITEGGGVYKANSKFIGITELSSCHGSEPCIHRGKPFGGCHGSEPCIHRGKPF-SSC 70
Db 61 GCGSGGS--KGGCGSGCGSCSCC---KPCCS-----SGC-GSSCCGSCCKPYCSQC 109

QY 71 HGSEPCIHRGKPFPGCHGS--EPCIHRGKPFSSCHGS--EPCIHRGKPFGGCHGSEP 126
Db 110 SCCRPPCSSSGRGSCCCKPCCSSSGCGSCCSCCKPC-----CSQSRC 159

QY 127 CI 128
Db 160 CV 161

RESULT 10
T31524
hypothetical protein Y116A8C.41 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T31524
R:McMurray, A.
submitted to the EMBL Data Library, October 1999
A:Reference number: T21041
A:Accession: T31524
A>Status: preliminary; translated from GB/EMBL/DBDJB
A:Molecule type: DNA
A:Residues: 1-232 <WIL>
A:Cross-references: EMBL:AL117204; PIDN:CAB55158.1; CESP:Y116A8C.41
A:Experimental source: clone Y116A8C
C:Genetics:
A:Gene: CESP:Y116A8C.41
A:Introns: 5/3; 57/3

Query Match 10.7%; Score 85; DB 2; Length 232;
Best Local Similarity 35.6%; Pred. No. 0.19;
Matches 36; Conservative 6; Mismatches 35; Indels 24; Gaps 10;

QY 45 HRGKPF-GGCHGSEPCIHRGKPF-SSCHGSEPCIHRGKPF-GGCHGSEPCI--IHRGK 99
Db 81 HRGETLRGTGTHRG--THRGTFRGTGTHRG--THRGTFRGTGTHRGTFGCTHGRGE 134

QY 100 PF-SSCHGSEPCIHRGKPF-----GGCHGSEPCIHRGK 133
Db 135 TFRSETHRG--THRGTFRGTGTHRG--THRGTFRGTGTHRGTFGCTHGRGE 169

RESULT 11
A38346
ultra-high-sulfur keratin 1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 24-Sep-1999
C:Accession: A38346
```

R;Wood, L.; Mills, M.; Hatzenbuehler, N.; Vogeli, G.  
J. Biol. Chem. 265, 21375-21380, 1990  
A:Title: Serine-rich ultra high sulfur protein gene expression in murine hair and skin  
A:Reference number: A38346; MUID:91065960  
A:Accession: A38346  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-230 <MOO>  
A:Cross-references: GB:M37759; NID:g200961; PIDN:AAA40106.1; PID:g200962  
C:Superfamily: ultra-high-sulfur keratin

Query Match 10.6%; Score 84; DB 2; Length 230;  
Best Local Similarity 36.0%; Pred. No. 0.24;  
Matches 41; Conservative 2; Mismatches 37; Indels 34; Gaps 12;

[illegible]

RESULT 12

12.  
 AL30051  
 T30173  
 zinc finger protein Pw1 - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
 C:Accession: T30173  
 R:Relaix, F.; Weng, X.; Marazzi, G.; Yang, E.; Copeland, N.; Jenkins, N.; Spence, S.E.;  
 Dev. Biol. 177, 383-396, 1996  
 A:Title: Pw1, a novel zinc finger gene implicated in the myogenic and neuronal lineages  
 A:Reference number: Z20754; MUID:96400442  
 A:Accession: T30173  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 1-1378 <REL>  
 A:Cross-references: EMBL:U48804; NID:g1197856; PID:g1197857; PIDN:AAC52770.1  
 A:Experimental source: strain sv129; limb bud  
 C:Genetics:  
 A:Map position: 7

Query Match 10.4%; Score 82.5; DB 2; Length 1378;  
Best Local Similarity 31.4%; Pred. No. 1.9;  
Matches 38; Conservative 8; Mismatches 28; Indels 47; Gaps 12;

Qy	37	HGSEPCIIHRGKPF	-----GCHGSEPC-----	IIHRGKPFSSCHGSEPCII	78
Db	754	HGQK---IHDEKPYGKEP	GKEPHGDEPQKEPLQEMRSEEP	HDDKP---HGQEP---	803
Qy	79	HRG-KPFGGCHGSEPCII	HRGKPFSSCHGSEPCII	HRGKPFG---GCHGSEPCII	HRGKPFSS
Db	804	HDDMRP---HGQEP---	HODEP-----HGQEP---	HGDEPHGQEPHGDEP---	HDKEPID
Qy	137	R	137		

0v 137 p 137

Db 847 0 847

RECEIVED 13

A72507  
 hypothetical protein APE2031 - Aeropyrum pernix (strain K1)  
 C:Species: Aeropyrum pernix  
 C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jun-2000  
 C:Accession: A72507  
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahashi, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kawarabayashi, T. 1999  
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix strain K1. 1999  
 A:Reference number: A72450; MUID:99310339

A;Accession: A72507  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-102 <RAW>  
A;Cross-references: DDBJ:AF000063; NID:G5105654; PIDN:BAA81041.1; PID:d1044827; PID:g  
A;Experimental source: strain K1  
C;Genetics:  
A;Gene: APE2031  
C;Superfamily: Aeropyrum pernix hypothetical protein APE2031

Query Match	10.3%	Score 82;	DB 2;	Length 102;
Best Local Similarity	32.3%	Pred. No. 0.17;		
Matches 30;	Conservative	7;	Mismatches 34;	Indels

Qy	52	GCHGSEPCIHRKPKPSSCHGSEPCIHRKPKPGCGHSGSEPCIHRKPKPS--SCHGSEP	109
		:  :	
Db	5	GCPGDGL-----RRRPGARIH-SRPC-----GGCSGR-CLRH7SRWHSDDGGCAGHTI	51
Qy	110	CIHRGKPFGGC-----HGSEPCIHRKPKPFSR	137
		:  :	
Db	52	QAGYAGSPFNGCCNHEMDSSRPRLQHPG--YSR	82

RESULT 14

T22674  
hypothetical protein F54F3.1 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #t  
C:Accession: T22674  
R:Percy, C.; Lloyd, C.  
submitted to the EMBL Data Library, September.1996  
A:Reference number: z19598  
A:Accession: T22674  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1584 <MIL>  
A:Cross-references: EMBL:z79696; PIDN:CAB01972.1; GSP  
A:Experimental source: clone F54F3  
C:Genetics:  
A:Gene: CESP:F54F3.1  
A:Map position: 5  
A:Introns: 35/3; 213/2; 157/3; 357/1; 440/3; 5

Query Match	10.3%	Score 81.5;	DB 2;	Length 1584;
Best Local Similarity	24.0%;	Pred. No. 2.7;		
Matches 29: Conservative	18:"	Mismatches	37:	Indels 37: Gaps 8:

Qy	32	ELSCSGSEPCIIHR-----GKPGGCGHSGEPCIIHRKPFSSCHGSEPC	76
Db	1068	ESSDCHGHCVNEHGAGEYICQLPFGSGDGFNCRGADQC--NP-SNP--SACYQNAHC	1134
Qy	77	I-----IHRGKPGGCGH--SEPCIIHRKPFSSCHGSEPCIIHRKPGFGGCHGSEPCII	129
Db	1125	VYDALLNAHAKCVDFGKDGTSVVPY--APATNCN-LEPRT-----CHANAOCVM	1172

0  
0  
130  
130

1173 1173 1173

DEC 15 1964

hypothetical protein RV3903c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: H70599  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
J.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,  
R.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

Query Match 10.2%; Score 81; DB 2; Length 846;  
Best Local Similarity 34.0%; Pred. No. 1.7;  
Matches 32; Conservative 9; Mismatches 33; Indels 20; Gaps 8;

QY 107 SEPCIIHRGKPFQ-GCH--GSEPCIIHRGKPFQ 137  
Db 634 EQQLAMHSNEPAGEGWHRLSDEAVDPQYGEPLSR 667

Search completed: April 14, 2001, 10:11:49  
Job time: 135 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2001, 10:16:54 ; Search time 41.8 Seconds  
(without alignments)  
6.434 Million cell updates/sec

Title: US-09-362-731-2

Perfect score: 79

Sequence: 1 HEIKKVLPGCHGS 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCrUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79	100.0	16	2	US-08-482-142-59
2	79	100.0	16	2	US-08-478-572-59
3	79	100.0	25	2	US-08-482-142-42
4	79	100.0	25	2	US-08-482-142-169
5	79	100.0	25	2	US-08-478-572-42
6	79	100.0	25	2	US-08-478-572-169
7	79	100.0	25	5	PCT-US95-04481-31
8	79	100.0	25	5	PCT-US95-04481-31
9	79	100.0	26	2	US-08-482-142-50
10	79	100.0	26	2	US-08-482-142-51
11	79	100.0	26	2	US-08-482-142-52
12	79	100.0	26	2	US-08-482-142-170
13	79	100.0	26	2	US-08-478-572-50
14	79	100.0	26	2	US-08-478-572-51
15	79	100.0	26	2	US-08-478-572-52
16	79	100.0	26	2	US-08-478-572-170
17	79	100.0	27	2	US-08-482-142-171
18	79	100.0	27	2	US-08-478-572-171
19	79	100.0	35	3	US-08-460-040-7
20	79	100.0	129	1	US-07-945-288-12
21	79	100.0	129	1	US-08-462-831-12
22	79	100.0	129	1	US-08-461-809-12
23	79	100.0	129	1	US-08-461-441-12
24	79	100.0	129	5	PCT-US93-08518-12
25	79	100.0	145	3	US-08-460-040-6
26	79	100.0	146	1	US-07-945-288-4
27	79	100.0	146	1	US-08-462-831-4

28 79 100.0 146 1 US-08-461-809-4 Sequence 4, Appli  
29 79 100.0 146 1 US-08-461-441-4 Sequence 4, Appli  
30 79 100.0 146 2 US-08-482-142-4 Sequence 4, Appli  
31 79 100.0 146 2 US-08-478-572-4 Sequence 4, Appli  
32 79 100.0 146 5 PCT-US93-08518-4 Sequence 4, Appli  
33 70 88.6 61 2 US-08-482-142-162 Sequence 162, App  
34 70 88.6 61 2 US-08-478-572-162 Sequence 162, App  
35 70 88.6 84 2 US-08-482-142-161 Sequence 161, App  
36 70 88.6 84 2 US-08-478-572-161 Sequence 161, App  
37 70 88.6 129 2 US-08-482-142-157 Sequence 157, App  
38 70 88.6 129 2 US-08-482-142-159 Sequence 159, App  
39 70 88.6 129 2 US-08-478-572-157 Sequence 157, App  
40 70 88.6 129 2 US-08-478-572-159 Sequence 159, App  
41 70 88.6 130 2 US-08-482-142-158 Sequence 158, App  
42 70 88.6 130 2 US-08-478-572-158 Sequence 158, App  
43 69 87.3 16 2 US-08-482-142-61 Sequence 61, Appl  
44 69 87.3 16 2 US-08-478-572-61 Sequence 61, Appl  
45 69 87.3 26 2 US-08-482-142-54 Sequence 54, Appl

#### ALIGNMENTS

RESULT 1  
US-08-482-142-59  
; Sequence 59, Application US/08482142  
; Patent No. 5820862  
; GENERAL INFORMATION:  
; APPLICANT: Garman, Richard  
; APPLICANT: Greenstein, Julia  
; APPLICANT: Kuo, Mei-chang  
; APPLICANT: Rogers, Bruce  
; APPLICANT: Franzen, Henry  
; APPLICANT: Chen, Xian  
; APPLICANT: Evans, Sean  
; APPLICANT: Shaked, Ze'ev  
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
; STREET: 610 LINCOLN STREET  
; CITY: WALTHAM  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII TEXT  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/482,142  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/445,307  
; FILING DATE: 07 June 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CRAIG, ANNE I.  
; REGISTRATION NUMBER: 32,976  
; REFERENCE/DOCKET NUMBER: 017,605  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 466-6000  
; TELEFAX: (617) 466-6040  
; INFORMATION FOR SEQ ID NO: 59:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 16 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: N-terminal  
US-08-482-142-59



Query Match 100.0%; Score 79; DB 2; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HEIKKVLVPGCHGS 14  
 |||||  
 DB 1 HEIKKVLVPGCHGS 14

## RESULT 2

US-08-478-572-59  
 ; Sequence 59, Application US/08478572  
 ; Patent No. 5968526  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Garman, Richard  
 ; APPLICANT: Greenstein, Julia  
 ; APPLICANT: Kuo, Mei-chang  
 ; APPLICANT: Rogers, Bruce  
 ; APPLICANT: Franzen, Henry  
 ; APPLICANT: Chen, Xian  
 ; APPLICANT: Evans, Sean  
 ; APPLICANT: Shaked, Ze'ev  
 ; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
 ; FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
 ; NUMBER OF SEQUENCES: 207  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
 ; STREET: 610 LINCOLN STREET  
 ; CITY: WALTHAM  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02154  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: ASCII TEXT  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/478,572  
 ; FILING DATE: 07-June-1995  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION NUMBER: 08/445,307  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: CRAIG, ANNE I.  
 ; REGISTRATION NUMBER: 32,976  
 ; REFERENCE/DOCKET NUMBER: 017.6US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617) 466-6000  
 ; TELEFAX: (617) 466-6040  
 ; INFORMATION FOR SEQ ID NO: 59:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 16 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; FRAGMENT TYPE: N-terminal  
 ; US-08-478-572-59

Query Match 100.0%; Score 79; DB 2; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HEIKKVLVPGCHGS 14  
 |||||  
 DB 1 HEIKKVLVPGCHGS 14

## RESULT 3

US-08-478-572-59

US-08-482-142-42  
 ; Sequence 42, Application US/08482142  
 ; Patent No. 5820862  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Garman, Richard  
 ; APPLICANT: Greenstein, Julia  
 ; APPLICANT: Kuo, Mei-chang  
 ; APPLICANT: Rogers, Bruce  
 ; APPLICANT: Franzen, Henry  
 ; APPLICANT: Chen, Xian  
 ; APPLICANT: Evans, Sean  
 ; APPLICANT: Shaked, Ze'ev  
 ; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
 ; FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
 ; NUMBER OF SEQUENCES: 207  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
 ; STREET: 610 LINCOLN STREET  
 ; CITY: WALTHAM  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02154  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: ASCII TEXT  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/482,142  
 ; FILING DATE: 07-JUN-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION NUMBER: US/08/445,307  
 ; FILING DATE: 07 June 1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: CRAIG, ANNE I.  
 ; REGISTRATION NUMBER: 32,976  
 ; REFERENCE/DOCKET NUMBER: 017.6US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617) 466-6000  
 ; TELEFAX: (617) 466-6040  
 ; INFORMATION FOR SEQ ID NO: 42:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 25 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; FRAGMENT TYPE: N-terminal  
 ; US-08-482-142-42

Query Match 100.0%; Score 79; DB 2; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HEIKKVLVPGCHGS 14  
 |||||  
 DB 1 HEIKKVLVPGCHGS 14

## RESULT 4

US-08-482-142-169  
 ; Sequence 169, Application US/08482142  
 ; Patent No. 5820862  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Garman, Richard  
 ; APPLICANT: Greenstein, Julia  
 ; APPLICANT: Kuo, Mei-chang  
 ; APPLICANT: Rogers, Bruce  
 ; APPLICANT: Franzen, Henry  
 ; APPLICANT: Chen, Xian  
 ; APPLICANT: Evans, Sean  
 ; APPLICANT: Shaked, Ze'ev

;; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
;; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)

;; NUMBER OF SEQUENCES: 207

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION

;; STREET: 610 LINCOLN STREET

;; CITY: WALTHAM

;; STATE: MA

;; COUNTRY: USA

;; ZIP: 02154

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: ASCII TEXT

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/482,142

;; FILING DATE: 07-JUN-1995

;; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US/08/445,307

;; FILING DATE: 07 June 1995

;; ATTORNEY/AGENT INFORMATION:

;; NAME: CRAIG, ANNE I.

;; REGISTRATION NUMBER: 32,976

;; REFERENCE/DOCKET NUMBER: 017.605

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (617) 466-6000

;; TELEFAX: (617) 466-6040

;; INFORMATION FOR SEQ ID NO: 169:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 25 amino acids

;; TYPE: amino acid

;; TOPOLOGY: linear

;; MOLECULE TYPE: peptide

;; FRAGMENT TYPE: Internal

;; US-08-482-142-169

Query Match 100.0%; Score 79; DB 2; Length 25;  
Best Local Similarity 100.0%; Pred. No. 3.3e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HEIKKVLVPGCHGS 14

|||||

Db 11 HEIKKVLVPGCHGS 24

RESULT 5

US-08-478-572-42

;; Sequence 42, Application US/08478572

;; Patent No. 5968526

;; GENERAL INFORMATION:

;; APPLICANT: Garman, Richard

;; APPLICANT: Greenstein, Julia

;; APPLICANT: Kuo, Mei-chang

;; APPLICANT: Rogers, Bruce

;; APPLICANT: Franzen, Henry

;; APPLICANT: Chen, Xian

;; APPLICANT: Evans, Sean

;; APPLICANT: Shaked, Ze'ev

;; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS

;; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)

;; NUMBER OF SEQUENCES: 207

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION

;; STREET: 610 LINCOLN STREET

;; CITY: WALTHAM

;; STATE: MA

;; COUNTRY: USA

;; ZIP: 02154

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: ASCII TEXT

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/478,572

;; FILING DATE: 07-June-1995

;; CLASSIFICATION:

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: 08/445,307

;; FILING DATE:

;; ATTORNEY/AGENT INFORMATION:

;; NAME: CRAIG, ANNE I.

;; REGISTRATION NUMBER: 32,976

;; REFERENCE/DOCKET NUMBER: 017.605

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (617) 466-6000

;; TELEFAX: (617) 466-6040

;; INFORMATION FOR SEQ ID NO: 42:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 25 amino acids

;; TYPE: amino acid

;; TOPOLOGY: linear

;; MOLECULE TYPE: peptide

;; FRAGMENT TYPE: N-terminal

;; US-08-478-572-42

Query Match 100.0%; Score 79; DB 2; Length 25;

Best Local Similarity 100.0%; Pred. No. 3.3e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HEIKKVLVPGCHGS 14

|||||

Db 1 HEIKKVLVPGCHGS 14

RESULT 6

US-08-478-572-169

;; Sequence 169, Application US/08478572

;; Patent No. 5968526

;; GENERAL INFORMATION:

;; APPLICANT: Garman, Richard

;; APPLICANT: Greenstein, Julia

;; APPLICANT: Kuo, Mei-chang

;; APPLICANT: Rogers, Bruce

;; APPLICANT: Franzen, Henry

;; APPLICANT: Chen, Xian

;; APPLICANT: Evans, Sean

;; APPLICANT: Shaked, Ze'ev

;; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS

;; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)

;; NUMBER OF SEQUENCES: 207

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION

;; STREET: 610 LINCOLN STREET

;; CITY: WALTHAM

;; STATE: MA

;; COUNTRY: USA

;; ZIP: 02154

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: ASCII TEXT

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/478,572

;; FILING DATE: 07-June-1995

;; CLASSIFICATION:

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: 08/445,307

;; FILING DATE:

;; ATTORNEY/AGENT INFORMATION:

;; NAME: CRAIG, ANNE I.



CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/445,307  
FILING DATE: 07 June 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: CRAIG, ANNE I.  
REGISTRATION NUMBER: 32,976  
REFERENCE/DOCKET NUMBER: 017.6US  
TELEPHONE: (617) 466-6000  
TELEFAX: (617) 466-6040  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-482-142-50

Query Match 100.0%; Score 79; DB 2; Length 26;  
Best Local Similarity 100.0%; Pred. No. 3.4e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HEIKKVLVPGCHGS 14  
Db 11 HEIKKVLVPGCHGS 24

RESULT 10  
US-08-482-142-51  
Sequence 51, Application US/08482142  
Patent No. 5820862  
GENERAL INFORMATION:  
APPLICANT: Garman, Richard  
APPLICANT: Greenstein, Julia  
APPLICANT: Kuo, Mei-chang  
APPLICANT: Rogers, Bruce  
APPLICANT: Franzen, Henry  
APPLICANT: Chen, Xian  
APPLICANT: Evans, Sean  
APPLICANT: Shaked, Ze'ev  
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
NUMBER OF SEQUENCES: 207  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
STREET: 610 LINCOLN STREET  
CITY: WALTHAM  
STATE: MA  
COUNTRY: USA  
ZIP: 02154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,142  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/445,307  
FILING DATE: 07 June 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: CRAIG, ANNE I.  
REGISTRATION NUMBER: 32,976  
REFERENCE/DOCKET NUMBER: 017.6US  
TELEPHONE: (617) 466-6000  
TELEFAX: (617) 466-6040  
INFORMATION FOR SEQ ID NO: 51:

SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-482-142-51

Query Match 100.0%; Score 79; DB 2; Length 26;  
Best Local Similarity 100.0%; Pred. No. 3.4e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HEIKKVLVPGCHGS 14  
Db 11 HEIKKVLVPGCHGS 24

RESULT 11  
US-08-482-142-52  
Sequence 52, Application US/08482142  
Patent No. 5820862  
GENERAL INFORMATION:  
APPLICANT: Garman, Richard  
APPLICANT: Greenstein, Julia  
APPLICANT: Kuo, Mei-chang  
APPLICANT: Rogers, Bruce  
APPLICANT: Franzen, Henry  
APPLICANT: Chen, Xian  
APPLICANT: Evans, Sean  
APPLICANT: Shaked, Ze'ev  
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
NUMBER OF SEQUENCES: 207  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
STREET: 610 LINCOLN STREET  
CITY: WALTHAM  
STATE: MA  
COUNTRY: USA  
ZIP: 02154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,142  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/445,307  
FILING DATE: 07 June 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: CRAIG, ANNE I.  
REGISTRATION NUMBER: 32,976  
REFERENCE/DOCKET NUMBER: 017.6US  
TELEPHONE: (617) 466-6000  
TELEFAX: (617) 466-6040  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-482-142-52

Query Match 100.0%; Score 79; DB 2; Length 26;  
Best Local Similarity 100.0%; Pred. No. 3.4e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HEIKKVLVPGCHGS 14  
 Db 11 HEIKKVLVPGCHGS 24

## RESULT 12

US-08-482-142-170  
 ; Sequence 170, Application US/08482142  
 ; Patent No. 5920862  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Garman, Richard  
 ; APPLICANT: Greenstein, Julia  
 ; APPLICANT: Kuo, Mei-chang  
 ; APPLICANT: Rogers, Bruce  
 ; APPLICANT: Franzen, Henry  
 ; APPLICANT: Chen, Xian  
 ; APPLICANT: Evans, Sean  
 ; APPLICANT: Shaked, Ze'ev  
 ; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
 ; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
 ; NUMBER OF SEQUENCES: 207  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
 ; STREET: 610 LINCOLN STREET  
 ; CITY: WALTHAM  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02154  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: ASCII TEXT  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/482,142  
 ; FILING DATE: 07-JUN-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/445,307  
 ; FILING DATE: 07 June 1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: CRAIG, ANNE I.  
 ; REGISTRATION NUMBER: 32,976  
 ; REFERENCE/DOCKET NUMBER: 017.6US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617) 466-6000  
 ; TELEFAX: (617) 466-6040  
 ; INFORMATION FOR SEQ ID NO: 170:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 26 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; FRAGMENT TYPE: internal  
 ; US-08-482-142-170

Qy 1 HEIKKVLVPGCHGS 14  
 Db 11 HEIKKVLVPGCHGS 24

## RESULT 13

US-08-478-572-50  
 ; Sequence 50, Application US/08478572  
 ; Patent No. 5968526  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Garman, Richard

; APPLICANT: Greenstein, Julia  
 ; APPLICANT: Kuo, Mei-chang  
 ; APPLICANT: Rogers, Bruce  
 ; APPLICANT: Franzen, Henry  
 ; APPLICANT: Chen, Xian  
 ; APPLICANT: Evans, Sean  
 ; APPLICANT: Shaked, Ze'ev  
 ; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
 ; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
 ; NUMBER OF SEQUENCES: 207  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
 ; STREET: 610 LINCOLN STREET  
 ; CITY: WALTHAM  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02154  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: ASCII TEXT  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/478,572  
 ; FILING DATE: 07-June-1995  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/445,307  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: CRAIG, ANNE I.  
 ; REGISTRATION NUMBER: 32,976  
 ; REFERENCE/DOCKET NUMBER: 017.6US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617) 466-6000  
 ; TELEFAX: (617) 466-6040  
 ; INFORMATION FOR SEQ ID NO: 50:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 26 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; FRAGMENT TYPE: N-terminal  
 ; US-08-478-572-50

Query Match 100.0%; Score 79; DB 2; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HEIKKVLVPGCHGS 14  
 Db 11 HEIKKVLVPGCHGS 24

## RESULT 14

US-08-478-572-51  
 ; Sequence 51, Application US/08478572  
 ; Patent No. 5968526  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Garman, Richard  
 ; APPLICANT: Greenstein, Julia  
 ; APPLICANT: Kuo, Mei-chang  
 ; APPLICANT: Rogers, Bruce  
 ; APPLICANT: Franzen, Henry  
 ; APPLICANT: Chen, Xian  
 ; APPLICANT: Evans, Sean  
 ; APPLICANT: Shaked, Ze'ev  
 ; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
 ; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
 ; NUMBER OF SEQUENCES: 207  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION

STREET: 610 LINCOLN STREET  
CITY: WALTHAM  
STATE: MA  
COUNTRY: USA  
ZIP: 02154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII TEXT  
CURRENT APPLICATION DATA:  
FILING DATE: 07-June-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/445,307  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: CRAIG, ANNE I.  
REGISTRATION NUMBER: 32,976  
REFERENCE/DOCKET NUMBER: 017.605  
TELEPHONE: (617) 466-6000  
TELEFAX: (617) 466-6040  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-478-572-51

Query Match 100.0%; Score 79; DB 2; Length 26;  
Best Local Similarity 100.0%; Pred. No. 3.4e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HEIKKVLVPGCHGS 14  
|||||  
Db 11 HEIKKVLVPGCHGS 24

## RESULT 15

US-08-478-572-52  
Sequence 52, Application US/08478572  
Patent No. 5968526  
GENERAL INFORMATION:  
APPLICANT: Garman, Richard  
APPLICANT: Greenstein, Julia  
APPLICANT: Kuo, Mei-chang  
APPLICANT: Rogers, Bruce  
APPLICANT: Franzen, Henry  
APPLICANT: Chen, Xian  
APPLICANT: Evans, Sean  
APPLICANT: Shaked, Ze'ev  
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS  
NUMBER OF SEQUENCES: 207  
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION  
STREET: 610 LINCOLN STREET  
CITY: WALTHAM  
STATE: MA  
COUNTRY: USA  
ZIP: 02154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/478,572

FILING DATE: 07-June-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/445,307  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: CRAIG, ANNE I.  
REGISTRATION NUMBER: 32,976  
REFERENCE/DOCKET NUMBER: 017.605  
TELEPHONE: (617) 466-6000  
TELEFAX: (617) 466-6040  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-478-572-52

Query Match 100.0%; Score 79; DB 2; Length 26;  
Best Local Similarity 100.0%; Pred. No. 3.4e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HEIKKVLVPGCHGS 14  
|||||  
Db 11 HEIKKVLVPGCHGS 24

Search completed: April 14, 2001, 10:16:55  
Job time: 440 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2001, 10:17:38 ; Search time 27.31 Seconds  
(without alignments)  
17.560 Million cell updates/sec

Title: US-09-362-731-2

Perfect score: 79

Sequence: 1 HEIKKVLVPGCHGS 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	79	100.0	146	1 DER2_DERPT	P49278 dermatophag
2	62	78.5	146	1 DEF2_DERFA	Q00855 dermatophag
3	42	53.2	424	1 OAT_YEAST	P07991 saccharomyc
4	41	51.9	397	1 YBP2_YEAST	P38226 saccharomyc
5	41	51.9	408	1 ARGD_BUCAI	P57600 buchera ap
6	41	51.9	1184	1 DP3A_MYCTU	Q10779 mycobacteri
7	40	50.6	257	1 YJE3_YEAST	P47053 saccharomyc
8	40	50.6	631	1 SUM1_LYTVA	Q00492 lytechinus
9	40	50.6	2499	1 ETFD_YEAST	Q08822 saccharomyc
10	40	50.6	2499	1 MPRI_BOVIN	P08169 bos taurus
11	39	49.4	440	1 LCAT_RAT	P18424 rattus norv
12	39	49.4	2483	1 MPRI_MOUSE	Q07113 mus musculu
13	38.5	48.7	726	1 TRF_BLABI	Q02942 blaberus di
14	38	48.1	188	1 LCAT_PIG	P30930 sus scrofa
15	38	48.1	342	1 TONB_PSEAE	Q51368 pseudomonas
16	38	48.1	438	1 LCAT_MOUSE	P16301 mus musculu
17	38	48.1	440	1 LCAT_HUMAN	P04180 homo sapien
18	38	48.1	440	1 LCAT_PAPAN	Q08758 papio anubi
19	38	48.1	440	1 LCAT_RABIT	P53761 oryctolagus
20	38	48.1	548	1 YDID_ECOLI	P38135 escherichia
21	38	48.1	725	1 MYTL_HUMAN	Q01538 homo sapien
22	38	48.1	803	1 FSPQ_XENLA	P35447 xenopus lae
23	38	48.1	912	1 NIA2_HORVU	P27969 hordeum vul
24	38	48.1	915	1 NIA1_HORVU	P27967 hordeum vul
25	38	48.1	1184	1 FBL2_HUMAN	P37889 homo sapien
26	38	48.1	1221	1 FBL2_MOUSE	P37889 mus musculu
27	37.5	47.5	379	1 CBID_SALTY	Q05628 salmonella
28	37	46.8	106	1 CYC3_DESSA	P00135 desulfovibr
29	37	46.8	123	1 ATPE_HELPJ	Q92k82 helicobacte
30	37	46.8	123	1 ATPE_HELPJ	P56084 helicobacte
31	37	46.8	287	1 TYSY_MYCPN	P78029 mycoplasma
32	37	46.8	301	1 FPG_RHINE	Q59752 rhizobium m
33	37	46.8	341	1 MURB_HAEIN	P44605 haemophilus

## RESULT 1

ID	DER2_DERPT	STANDARD;	PRT;	146 AA.
AC	P49278;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	MITE ALLERGEN DER P 2 PRECURSOR (DER P II) (DPX).			
GN	DERP2.			
OS	Dermatophagoides pteronyssinus (House-dust mite).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;			
OC	Acariformes; Sarcopitiformes; Astigmata; Analgoidea; Pyroglyphidae;			
OC	Dermatophagoides.			
OX	NCBI_TaxID=6956;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90256301; PubMed=2341191;			
RA	Chua K.Y., Doyle C.R., Simpson R.J., Turner K.J., Stewart G.A.,			
RA	Thomas W.R.;			
RT	"Isolation of cDNA coding for the major mite allergen Der p II by IgE			
RT	plaque immunoassay."			
RL	Int. Arch. Allergy Appl. Immunol. 91:118-123(1990).			
RN	[2]			
RP	PARTIAL SEQUENCE OF 18-57.			
RX	MEDLINE=89278484; PubMed=2732406;			
RA	Heymann P.W., Chapman M.D., Aalberse R.C., Fox J.W.,			
RA	Platts-Mills T.A.;			
RT	"Antigenic and structural analysis of group II allergens (Der f II			
RT	and Der p II) from house dust mites (Dermatophagoides spp).";			
RL	J. Allergy Clin. Immunol. 83:1055-1067(1989).			
RN	[3]			
RP	VARIANTS.			
RA	Smith W., Hales B.J., Thomas W.R.;			
RT	"Allergens of wild house dust mites: environmental Der p I and Der p 2			
RT	sequence polymorphisms."			
RL	Submitted (JUN-2000) to the SWISS-PROT data bank.			
RN	[4]			
RP	STRUCTURE BY NMR.			
RX	MEDLINE=98409423; PubMed=9737847;			
RA	Mueller G.A., Benjamin D.C., Rule G.S.;			
RT	"Tertiary structure of the major house dust mite allergen Der p 2:			
RT	sequential and structural homologies."			
RL	Biochemistry 37:12707-12714(1998).			
CC	1- SIMILARITY: TO MITE ALLERGEN LEP D I.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement. See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch.			
CC	-----			
DR	EMBL: AF276239; AAF86462.1;			
DR	PDB: 1A9V; 14-OCT-98.			
KW	Allergen; Signal; 3D-structure; Polymorphism.			

P54895 streptomyces  
P03338 feline sarc  
O29685 archaeoglob  
P57473 buchera ap  
O81395 zea mays (m  
P10262 feline leuk  
Q9Y4K1 homo sapien  
P00336 desulfovibr  
O57580 methanococc  
O28355 archaeoglob  
P11044 bacillus su  
P38878 saccharomyc

## ALIGNMENTS



```

FT SIGNAL 1 17
FT CHAIN 18 146
FT DISULFID 25 136
FT DISULFID 38 44
FT DISULFID 90 95
FT VARIANT 39 39
FT VARIANT 40 40
FT VARIANT 44 44
FT VARIANT 47 47
FT VARIANT 49 49
FT VARIANT 56 56
FT VARIANT 61 61
FT VARIANT 75 75
FT VARIANT 78 78
FT VARIANT 81 81
FT VARIANT 95 95
FT VARIANT 98 98
FT VARIANT 108 108
FT VARIANT 111 111
FT VARIANT 114 114
FT VARIANT 116 116
FT VARIANT 118 118
FT VARIANT 127 127
SQ SEQUENCE 146 AA; 15999 MW; 591B2FA7FD26D3AF CRC64;

Query Match 100.0%; Score 79; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 7.2e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HEIKKVLVPGCHGS 14
Db 28 HEIKKVLVPGCHGS 41

RESULT 2
DEF2_DERFA STANDARD; PRT; 146 AA.
AC Q00855; P39672; Q26359;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MITE ALLERGEN DER F 2 PRECURSOR (DER F II).
GN DERF2.
OS Dermatophagoides farinae (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcopiformes; Astigmata; Analgoidea; Pyroglyphidae;
OC Dermatophagoides.
OX NCBI_TaxID=6954;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=91291341; PubMed=1368682;
RA Yuuki T., Okumura Y., Ando T., Yamakawa H., Suko M., Haida M.,
RA Okudaira H.;
RT "Cloning and expression of cDNA coding for the major house dust mite
RT allergen Der f II in Escherichia coli.";
RL Agric. Biol. Chem. 55:1233-1238(1991).
RN [2]
RP SEQUENCE OF 4-146 FROM N.A.
RX MEDLINE=94256850; PubMed=8198452;
RA Okuhira H.;
RT "Molecular biology of mite antigens.";
RL Arerugi 43:435-440(1994).
RN [3]
RP DISULFIDE BONDS, AND PARTIAL SEQUENCE.
RX MEDLINE=93283958; PubMed=8508052;
RA Nishiyama C., Yuuki T., Takai T., Okumura Y., Okudaira H.;
RT "Determination of three disulfide bonds in a major house dust mite
RT allergen, Der f II.";
RL Int. Arch. Allergy Immunol. 101:159-166(1993).
RN [4]
RP PARTIAL SEQUENCE OF 18-52.
RX MEDLINE=9278484; PubMed=2732406;

```

```

RA Heymann P.W., Chapman M.D., Aalberse R.C., Fox J.W.,
RA Platts-Mills T.A.;
RT "Antigenic and structural analysis of group II allergens (Der f II
RT and Der p II) from house dust mites (Dermatophagoides spp).";
RL J. Allergy Clin. Immunol. 83:1055-1067(1989).
RN [5]
RP STRUCTURE BY NMR.
RX MEDLINE=98079088; PubMed=9417088;
RA Ichikawa S., Hatanaka H., Yuuki T., Iwamoto N., Kojima S.,
RA Nishiyama C., Ogura K., Okumura Y., Inagaki F.;
RT "Solution structure of Der f 2, the major mite allergen for atopic
RT diseases.";
RL J. Biol. Chem. 273:356-360(1998).
CC -I- MISCELLANEOUS: THE SEQUENCE SHOWN HERE IS FROM CLONE 2. THE
CC N-TERMINAL SEQUENCE (AA 1-8) FROM CLONE 1 AND 11 ARE NOT YET
CC KNOWN.
CC -I- SIMILARITY: TO MITE ALLERGEN LEP D I.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D10447; BAA01239.1; -
CC EMBL; D10448; BAA01240.1; -
CC EMBL; D10449; BAA01241.1; -
CC EMBL; S70378; AAB30829.1; -
CC PIR; PS0417; PS0417.
CC PDB; 1AHK; 08-APR-98.
CC PDB; 1AHM; 08-APR-98.
KW Allergen; Signal; Polymorphism; 3D-structure.
FT SIGNAL 1 17
FT CHAIN 18 146 MITE ALLERGEN DER F 2.
FT DISULFID 25 136
FT DISULFID 38 44
FT DISULFID 90 95
FT VARIANT 93 93 M -> V (IN CLONE 1).
FT VARIANT 105 105 I -> A (IN CLONE 11).
FT VARIANT 128 128 I -> V (IN CLONE 11).
FT VARIANT 142 142 G -> A (IN CLONE 11).
FT CONFLICT 5 8 ILCL -> GTMV (IN REF. 2).
SQ SEQUENCE 146 AA; 15802 MW; FA118206CD88534A CRC64;

Query Match 78.5%; Score 62; DB 1; Length 146;
Best Local Similarity 78.6%; Pred. No. 0.00069;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HEIKKVLVPGCHGS 14
Db 28 HEIKKVLVPGCHGS 41

RESULT 3
OAT_YEAST
ID OAT_YEAST STANDARD; PRT; 424 AA.
AC P07991;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE ORNITHINE AMINOTRANSFERASE (EC 2.6.1.13) (ORNITHINE--OXO-ACID
DE AMINOTRANSFERASE).
GN CAR2 OR CARGB OR YLR438W OR L9753.2.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88055042; PubMed=2824201;

```



```

RT RT "Genome sequence of the endocellular bacterial symbiont of aphids
RL Buchnera sp. APS.",
RL Nature 407:81-86(2000).
CC -!- CATALYTIC ACTIVITY: N2-ACETYL-L-ORNITHINE + 2-OXOGLUTARATE =
CC N-ACETYL-L-GLUTAMATE 5-SEMIALDEHYDE + L-GLUTAMATE.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -!- PATHWAY: FOURTH STEP IN ARGININE BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AP001119; BAB13227.1; -
CC DR PROSITE; PS00600; AA_TRANSFER_CLASS_3; 1.
CC DR Transferrase; Aminotransferase; Pyridoxal phosphate;
CC KW Arginine biosynthesis.
CC FT BINDING 255 255 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC FT SEQUENCE 408 AA; 45295 MW; 5278FD811211F419 CRC64;
CC SQ
CC -----
Query Match 51.9%; Score 41; DB 1; Length 408;
Best Local Similarity 42.9%; Pred. No. 8.9;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 1 HEKKKVLVPCGHS 14
:|: |::|||:
DB 270 NEVASVIAPGINGT 283
CC -----
RESULT 6
DP3A_MYCTU
ID DP3A_MYCTU STANDARD; PRT; 1184 AA.
IC Q10779;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROBABLE DNA POLYMERASE III, ALPHA CHAIN (EC 2.7.7.7).
DN DNAE OR DNAEI OR RV1547 OR MTCY48.18C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekla F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrall B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -----
CC -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC THE ALPHA CHAIN IS THE DNA POLYMERASE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
CC N PYROPHOSPHATE + DNA(N).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Echinoidea; Euechinoidea; Echinacea; Temnopleuroidea; Toxopneustidae;  
 OC Lytechinus.  
 OX NCBI\_TaxID=7654;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91296792; PubMed=2068103;  
 RA Venuti J.M., Goldberg L., Chakraborty T., Olson E.N., Klein W.H.;  
 RT "A myogenic factor from sea urchin embryos capable of programming  
 muscle differentiation in mammalian cells.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:6219-6223(1991).  
 CC -!- FUNCTION: REGULATORY FACTOR DURING EMBRYOGENESIS. CONVERSION  
 CC OF PLURIPOTENT SECONDARY MESENCHYME CELLS TO MYOGENIC CELLS.  
 CC IT BINDS TO THE MCK ENHANCER ELEMENT.  
 CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. HOMODIMER, AND HETERODIMER WITH THE UBIQUITOUS BHLH  
 CC PROTEIN E12.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED UP FROM THE GASTRULATION STAGE,  
 CC MAXIMAL CONCENTRATION DURING THE PRISM STAGE AND DIMINISHING  
 CC CONCENTRATIONS IN THE DIFFERENTIATING MYOCYTES.  
 CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. "MYOGENIC FACTORS" SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M69052; AAA30009.1; ALT\_INIT.  
 DR PIR; A41123; A41123.  
 DR HSP; P10085; LMDY.  
 DR TRANSFAC; T00926; -.  
 DR InterPro; IPR001092; -.  
 DR InterPro; IPR002546; -.  
 DR InterPro; IPR003015; -.  
 DR Pfam; PF01586; Basic; 1.  
 DR Pfam; PF00010; HLH; 1.  
 DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; 1.  
 KW Transcription regulation; DNA-binding; Nuclear protein.  
 FT DNA\_BIND 112 124 BASIC DOMAIN.  
 FT DOMAIN 125 164 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT SEQUENCE 260 AA; 28682 MW; 8CC5FA9D2E6DAFA4 CRC64;  
 SQ  
 Query Match 50.6%; Score 40; DB 1; Length 260;  
 Best Local Similarity 58.3%; Pred. No. 8.7;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 ELKKVLVPGCHG 13  
 I::| | | | |  
 Db 79 ELEHVLAPGFHG 90  
 RESULT 9  
 ID ETFD\_YEAST STANDARD; PRT; 631 AA.  
 AC Q08822;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE PROBABLE ELECTRON TRANSFER FLAVOPROTEIN-UBIQUINONE OXIDOREDUCTASE  
 DE PRECURSOR (EC 1.5.5.1) (ETF-OO) (ETF-UBIQUINONE OXIDOREDUCTASE) (ETF  
 DE DEHYDROGENASE) (ELECTRON-TRANSFERRING-FLAVOPROTEIN DEHYDROGENASE).  
 GN YOR356W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RA Delius H., Hebling U., Hofmann B.;  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: ACCEPTS ELECTRONS FROM ETF AND REDUCES UBIQUINONE (BY  
 CC SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: REDUCED ETF + UBIQUINONE -> ETF + UBIQUINOL.  
 CC -!- COFACTOR: FAD AND A 4FE-4S CLUSTER (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE (BY  
 CC SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE ETF-OO / FIXC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z75264; CAA99685.1; -.  
 DR SGD; S0005883; YOR356W.  
 KW Oxidoreductase; Electron transport; Flavoprotein; FAD; Iron-sulfur;  
 KW 4Fe-4S; Mitochondrion; Transit peptide; Ubiquinone  
 FT TRANSIT 1 631 ? MITOCHONDRION (POTENTIAL).  
 FT CHAIN 65 79 PROBABLE ELECTRON TRANSFER FLAVOPROTEIN-  
 FT NP\_BIND 65 79 UBIQUINONE OXIDOREDUCTASE.  
 FT METAL 574 574 FAD (ADP PART) (POTENTIAL).  
 FT METAL 600 600 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 603 603 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 606 606 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 SQ SEQUENCE 631 AA; 69634 MW; 7493FG7093D88391 CRC64;  
 Query Match 50.6%; Score 40; DB 1; Length 631;  
 Best Local Similarity 77.8%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 6 VLVPGCHGS 14  
 I::| | | | |  
 Db 242 VLAEGCHGS 250  
 RESULT 10  
 MPRI\_BOVIN STANDARD; PRT; 2499 AA.  
 ID MPRI\_BOVIN STANDARD; PRT; 2499 AA.  
 AC P08169;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE CATION-INDEPENDENT MANNOSE-6-PHOSPHATE RECEPTOR PRECURSOR (CI MAN-6-P  
 DE RECEPTOR) (CI-MPR) (INSULIN-LIKE GROWTH FACTOR II RECEPTOR) (300 KDA  
 DE MANNOSE 6-PHOSPHATE RECEPTOR) (MPR 300) (MPR300).  
 GN IGF2R.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88115411; PubMed=2963004;  
 RA Lobel P., Dahms N.M., Kornfeld S.;  
 RT "Cloning and sequence analysis of the cation-independent mannose 6-  
 RT phosphate receptor.";  
 RL J. Biol. Chem. 263:2563-2570(1988).  
 CC [2]  
 RP SEQUENCE OF 1039-2499 FROM N.A.  
 RX MEDLINE=87175648; PubMed=2951738;  
 RA Lobel P., Dahms N.M., Breitmeyer J., Chirgwin J.M., Kornfeld S.;  
 RT "Cloning of the bovine 215-kDa cation-independent mannose 6-phosphate  
 RT receptor.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:2233-2237(1987).



Best Local Similarity 50.0%; Pred. No. 22;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 HEIKKVLVPGCHGS 14  
| :|||||I:  
Db 45 HTRPVLVPGCMGN 58

RESULT 12  
MPRI\_MOUSE STANDARD; PRT; 2483 AA.  
ID MPRI\_MOUSE Q61822;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE CATION-INDEPENDENT MANNOSE-6-PHOSPHATE RECEPTOR PRECURSOR (CI MAN-6-P  
DE RECEPTOR) (CI-MPR) (INSULIN-LIKE GROWTH FACTOR II RECEPTOR) (300 KDA  
DE MANNOSE 6-PHOSPHATE RECEPTOR) (MPR 300) (WPR300).  
GN IGF2R.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94245146; PubMed=8188212;  
RA Szebenyi G., Rotwein P.;  
RT "The mouse insulin-like growth factor II/cation-independent mannos  
RT 6-phosphate (IGF-II/MPR) receptor gene: molecular cloning and  
RT genomic organization.";  
RL Genomics 19:120-129(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=C57BL/6 X CBA; TISSUE=Liver;  
RX MEDLINE=94252588; PubMed=8194771;  
RA Ludwig T., Tenschler K., Remmler J., Hoflack B., Lobel P.;  
RT "Cloning and sequencing of cDNAs encoding the full-length mouse  
RT mannos 6-phosphate (IGF-II/MPR) receptor gene.";  
RL Gene 142:311-312(1994).  
RN [3]  
RP SEQUENCE OF 1-44 AND 93-106 FROM N.A.  
RX STRAIN=C57BL/6, AND 129;  
RX MEDLINE=93214996; PubMed=8462104;  
RA Stoger R., Kubicka P., Liu C.G., Kafri T., Razin A., Cedar H.,  
RA Barlow D.P.;  
RT "Maternal-specific methylation of the imprinted mouse Igf2r locus  
RT identifies the expressed locus as carrying the imprinting signal.";  
RL Cell 73:61-71(1993).  
RN [4]  
RP SEQUENCE OF 1-44 FROM N.A.  
RX STRAIN=129/SV;  
RX MEDLINE=96130821; PubMed=8584025;  
RA Liu Z., Mittanck D.W., Kim S., Rotwein P.;  
RT "Control of insulin-like growth factor-II/mannose 6-phosphate  
RT receptor gene transcription by proximal promoter elements.";  
RL Mol. Endocrinol. 9:1477-1487(1995).  
RN [5]  
RP SEQUENCE OF 435-488 FROM N.A.  
RX TISSUE=Liver;  
RX MEDLINE=91170218; PubMed=1848553;  
RA Szebenyi G., Rotwein P.;  
RT "Differential regulation of mannos 6-phosphate receptors and their  
RT ligands during the myogenic development of C2 cells.";  
RL J. Biol. Chem. 266:5534-5539(1991).  
RN [6]  
RP SEQUENCE OF 1625-2045 FROM N.A.  
RX STRAIN=C57BL/6;  
RA Matzner U.;  
RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: TRANSPORT OF PHOSPHORYLATED LYSOSOMAL ENZYMES FROM  
CC THE GOLGI COMPLEX AND THE CELL SURFACE TO LYSOSOMES. LYSOSOMAL  
CC ENZYMES BEARING PHOSPHOMANNOSYL RESIDUES BIND SPECIFICALLY TO  
CC MANNOSE-6-PHOSPHATE RECEPTORS IN THE GOLGI APPARATUS AND THE

CC RESULTING RECEPTOR-LIGAND COMPLEX IS TRANSPORTED TO AN ACIDIC  
CC PRELYSOSOMAL COMPARTMENT WHERE THE LOW PH MEDIATES THE DISSOCIATION  
CC OF THE COMPLEX. THIS RECEPTOR ALSO BINDS INSULIN GROWTH FACTOR II.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. LYSOSOMAL.  
CC -1- DOMAIN: CONTAINS 15 REPEATING UNITS OF APPROXIMATELY 147 AA. THE  
CC MOST HIGHLY CONSERVED REGION WITHIN THE REPEAT CONSISTS OF A  
CC STRETCH OF 13 AA THAT CONTAINS CYSTEINES AT BOTH ENDS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L22143; AAA39320.1; JOINED.  
DR EMBL; L22096; AAA39320.1; JOINED.  
DR EMBL; L22097; AAA39320.1; JOINED.  
DR EMBL; L22098; AAA39320.1; JOINED.  
DR EMBL; L22099; AAA39320.1; JOINED.  
DR EMBL; L22100; AAA39320.1; JOINED.  
DR EMBL; L22101; AAA39320.1; JOINED.  
DR EMBL; L22102; AAA39320.1; JOINED.  
DR EMBL; L22103; AAA39320.1; JOINED.  
DR EMBL; L22104; AAA39320.1; JOINED.  
DR EMBL; L22105; AAA39320.1; JOINED.  
DR EMBL; L22106; AAA39320.1; JOINED.  
DR EMBL; L22107; AAA39320.1; JOINED.  
DR EMBL; L22108; AAA39320.1; JOINED.  
DR EMBL; L22109; AAA39320.1; JOINED.  
DR EMBL; L22110; AAA39320.1; JOINED.  
DR EMBL; L22111; AAA39320.1; JOINED.  
DR EMBL; L22112; AAA39320.1; JOINED.  
DR EMBL; L22113; AAA39320.1; JOINED.  
DR EMBL; L22114; AAA39320.1; JOINED.  
DR EMBL; L22115; AAA39320.1; JOINED.  
DR EMBL; L22116; AAA39320.1; JOINED.  
DR EMBL; L22117; AAA39320.1; JOINED.  
DR EMBL; L22118; AAA39320.1; JOINED.  
DR EMBL; L22119; AAA39320.1; JOINED.  
DR EMBL; L22120; AAA39320.1; JOINED.  
DR EMBL; L22121; AAA39320.1; JOINED.  
DR EMBL; L22122; AAA39320.1; JOINED.  
DR EMBL; L22123; AAA39320.1; JOINED.  
DR EMBL; L22124; AAA39320.1; JOINED.  
DR EMBL; L22125; AAA39320.1; JOINED.  
DR EMBL; L22126; AAA39320.1; JOINED.  
DR EMBL; L22127; AAA39320.1; JOINED.  
DR EMBL; L22128; AAA39320.1; JOINED.  
DR EMBL; L22129; AAA39320.1; JOINED.  
DR EMBL; L22130; AAA39320.1; JOINED.  
DR EMBL; L22131; AAA39320.1; JOINED.  
DR EMBL; L22132; AAA39320.1; JOINED.  
DR EMBL; L22133; AAA39320.1; JOINED.  
DR EMBL; L22134; AAA39320.1; JOINED.  
DR EMBL; L22135; AAA39320.1; JOINED.  
DR EMBL; L22136; AAA39320.1; JOINED.  
DR EMBL; L22137; AAA39320.1; JOINED.  
DR EMBL; L22138; AAA39320.1; JOINED.  
DR EMBL; L22139; AAA39320.1; JOINED.  
DR EMBL; L22140; AAA39320.1; JOINED.  
DR EMBL; L22141; AAA39320.1; JOINED.  
DR EMBL; L22142; AAA39320.1; JOINED.  
DR EMBL; U04710; AAA19568.1; -  
DR EMBL; L06445; AAA37921.1; -  
DR EMBL; L06446; AAA37922.1; -  
DR EMBL; U26348; AAA98844.1; -  
DR EMBL; M58586; AAA39483.1; -  
DR EMBL; X60389; CAA42940.1; -  
DR HSSP; P02784; 1PDC.  
DR MGD; MGI:96435; Igf2r.  
DR InterPro; IPR000479; -.



Query Match 48.7%; Score 38.5; DB 1; Length 726;  
 Best Local Similarity 52.9%; Pred. No. 43;  
 Matches 9; Conservative 0; Mismatches 3; Indels 5; Gaps 1;

QY 1 HEIKKVLVP-----GCH 12  
 DB 26 HEIKKVCVPEGALESCH 42

## RESULT 14

LCAT\_PIG  
 ID LCAT\_PIG STANDARD; PRT; 188 AA.  
 AC P30930;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE PHOSPHATIDYLCHOLINE-STEROL ACYLTRANSFERASE (EC 2.3.1.43) (LECITHIN-  
 CHOLESTEROL ACYLTRANSFERASE) (PHOSPHOLIPID-CHOLESTEROL  
 DE ACYLTRANSFERASE) (FRAGMENTS).  
 GN LCAT.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Plasma;  
 RX MEDLINE=90075705; PubMed=2591200;  
 RA Yueskel K.U., Park Y.B., Jung J., Gracy R.W., Lacko A.G.;  
 RT "Studies on the structure of lecithin:cholesterol acyltransferase  
 (LACT) -- comparisons of the active site region and secondary  
 structure of the human and the porcine enzymes.";  
 RT Comp. Biochem. Physiol. 94B:389-394(1989).  
 RN [2]  
 RP SEQUENCE OF 97-106 AND 142-154.  
 RC TISSUE=Plasma;  
 RX MEDLINE=87156771; PubMed=3827927;  
 RA Park Y.B., Yueskel K.U., Gracy R.W., Lacko A.G.;  
 RT "The catalytic center of lecithin:cholesterol acyltransferase:  
 RT isolation and sequence of diisopropyl fluorophosphate-labeled  
 RT peptides";  
 RL Biochem. Biophys. Res. Commun. 143:360-363(1987).  
 CC -!- FUNCTION: CENTRAL ENZYME IN THE EXTRACELLULAR METABOLISM OF PLASMA  
 CC LIPOPROTEINS. AMONG OTHER SUBSTRATES IT ESTERIFIES THE FREE  
 CC CHOLESTEROL TRANSPORTED IN PLASMA LIPOPROTEINS.  
 CC -!- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + STEROL = STEROL ESTER +  
 CC 1-ACYLGLYCEROPHOSPHOCHOLINE (PALMITOYL, OLEOYL, AND LINOLEOYL CAN  
 CC BE TRANSFERRED; A NUMBER OF STEROLS, INCLUDING CHOLESTEROL, CAN  
 CC ACT AS ACCEPTOR).  
 CC -!- ENZYME REGULATION: APOLIPOPROTEIN A-I IS A POTENT ACTIVATOR FOR  
 CC THIS ENZYME.  
 CC -!- SIMILARITY: PARTIAL WITH LIPASES (PANCREATIC, GASTRIC, HEPATIC,  
 CC LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).  
 DR PIR; P0153; P0153.  
 DR PIR; A29544; A29544.  
 DR PIR; B29544; B29544.  
 DR InterPro: IPR000734.  
 DR PROSITE: PS00120; LIPASE\_SER; PARTIAL.  
 KW Transferase; Acyltransferase; Lipid metabolism; Glycoprotein.  
 FT CARBOHYD 20 20 N-LINKED (GLCNAC. . .).  
 FT NON\_CONS 34 35  
 FT NON\_CONS 44 45  
 FT NON\_CONS 60 61  
 FT NON\_CONS 66 67  
 FT NON\_CONS 77 78  
 FT NON\_CONS 84 85  
 FT NON\_CONS 96 97  
 FT NON\_CONS 106 107  
 FT NON\_CONS 115 116  
 FT NON\_CONS 154 155  
 FT NON\_TER 188 188  
 SQ SEQUENCE 188 AA; 21232 MW; A2A25EDB015EAB48 CRC64;

Query Match 48.1%; Score 38; DB 1; Length 188;  
 Best Local Similarity 50.0%; Pred. No. 14;  
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 HEIKKVLVPGCHGS 14  
 DB 21 HTRPVILVPGCLGN 34

## RESULT 15

TONB\_PSEAE  
 ID TONB\_PSEAE STANDARD; PRT; 342 AA.  
 AC Q51368;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE TONB PROTEIN.  
 GN TONB OR PA5531.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=96262716; PubMed=8704984;  
 RA Poole K., Zhao Q., Neshat S., Heinrichs D.E., Dean C.R.;  
 RT "The Pseudomonas aeruginosa tonB gene encodes a novel TonB protein.";  
 RL Microbiology 142:1449-1458(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 CC -!- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT  
 CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO  
 CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO  
 CC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-  
 CC REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE  
 CC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER  
 CC MEMBRANE PROTEINS (BY SIMILARITY).  
 CC -!- SUBUNIT: THE ACCESSORY PROTEINS EXBB AND EXBD SEEM TO FORM A  
 CC COMPLEX WITH TONB (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: PERIPLASMIC. ANCHORED TO THE CYTOPLASMIC  
 CC MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE  
 CC PERIPLASM (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE TONB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC ENBL; U23764; AAB18654.1;  
 DR ENBL; AE004965; AAG08916.1;  
 KW Transport; Protein transport; Inner membrane; Periplasmic;  
 KW Transmembrane; Signal-anchor; Repeat.  
 FT DOMAIN 1 87 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 88 108 SIGNAL-ANCHOR (POTENTIAL).  
 FT DOMAIN 109 342 PERIPLASMIC (POTENTIAL).  
 FT DOMAIN 159 181 9 X 2 REPEATS OF K-P.



FT DOMAIN 190 199 4 X 2 REPEATS OF K-P.  
FT CONFLICT 272 272 G -> R (IN REF. 1).  
SQ SEQUENCE 342 AA; 36892 MW; 157F036B39E9152D CRC64;

Query Match 48.1%; Score 38; DB 1; Length 342;  
Best Local Similarity 50.0%; Pred. No. 25;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 3 IKKVLVPGCHGS 14  
Db :::::|  
62 VEEVLIPYAHGS 73

Search completed: April 14, 2001, 10:17:40  
Job time: 342 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2001, 10:13:06 ; Search time 77.31 Seconds  
(without alignments)  
21.225 Million cell updates/sec

Title: US-09-362-731-2

Perfect score: 79

Sequence: 1 HEIKKLVPGCHGS 14

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_15.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_unclassified.\*
- 13: sp\_vertebrate.\*
- 14: sp\_virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	68	86.1	135	5 Q9TZZ2	Q9TZZ2 euroglyphus
2	68	86.1	145	5 Q96430	Q96430 euroglyphus
3	44	55.7	292	10 Q9ZU90	Q9ZU90 arabidopsis
4	43.5	55.1	1083	4 Q9UQ06	Q9UQ06 homo sapien
5	43.5	55.1	1117	4 Q9ULD8	Q9ULD8 homo sapien
6	43	54.4	226	4 Q9UJG9	Q9UJG9 homo sapien
7	43	54.4	382	10 Q9SIM7	Q9SIM7 arabidopsis
8	43	54.4	917	4 Q9UFU8	Q9UFU8 homo sapien
9	42	53.2	697	2 Q45419	Q45419 bacillus st
10	41	51.9	276	11 Q63020	Q63020 rattus norv
11	41	51.9	288	2 Q9PNM5	Q9PNM5 campylobact
12	41	51.9	475	5 Q46038	Q46038 drosophila
13	41	51.9	491	5 Q9W508	Q9W508 drosophila
14	41	51.9	573	1 Q58294	Q58294 pyrococcus
15	41	51.9	1631	4 Q9Y6U6	Q9Y6U6 homo sapien
16	40	50.6	205	2 Q9RPQ8	Q9RPQ8 vibrio chol
17	40	50.6	224	4 Q9UJF5	Q9UJF5 homo sapien
18	40	50.6	230	4 Q95553	Q95553 homo sapien
19	40	50.6	246	4 Q95552	Q95552 homo sapien

20	40	50.6	271	11 Q9R2A8	Q9R2A8 mus musculu
21	40	50.6	315	4 Q9P0X8	Q9P0X8 homo sapien
22	40	50.6	319	11 Q9QUH4	Q9QUH4 mus musculu
23	40	50.6	321	4 Q9P0Y1	Q9P0Y1 homo sapien
24	40	50.6	325	11 Q61110	Q61110 mus musculu
25	40	50.6	337	4 Q9P0X7	Q9P0X7 homo sapien
26	40	50.6	338	11 Q88972	Q88972 mus musculu
27	40	50.6	341	4 Q9P0X6	Q9P0X6 homo sapien
28	40	50.6	341	11 Q9QYS9	Q9QYS9 mus musculu
29	40	50.6	344	11 Q9Z246	Q9Z246 mus musculu
30	40	50.6	347	4 Q9P0Y0	Q9P0Y0 homo sapien
31	40	50.6	363	4 Q9P0X9	Q9P0X9 homo sapien
32	40	50.6	383	13 Q42476	Q42476 brachydanio
33	40	50.6	438	10 Q9MAH3	Q9MAH3 pseudomonas
34	40	50.6	504	2 P72236	P72236 pseudomonas
35	40	50.6	1146	3 Q06685	Q06685 saccharomyc
36	39.5	50.0	1087	11 Q9WVJ0	Q9WVJ0 mus musculu
37	39.5	50.0	1087	11 Q89047	Q89047 rattus norv
38	39	49.4	235	10 Q9SMA5	Q9SMA5 oryza sativ
39	39	49.4	325	2 Q9K7S6	Q9K7S6 bacillus ha
40	39	49.4	332	13 Q9YH19	Q9YH19 gallus gall
41	39	49.4	340	13 Q9YH20	Q9YH20 gallus gall
42	39	49.4	349	13 Q9YH18	Q9YH18 gallus gall
43	39	49.4	440	11 Q35849	Q35849 rattus norv
44	39	49.4	578	10 Q9M8N0	Q9M8N0 arabidopsis
45	39	49.4	670	10 Q9LV04	Q9LV04 arabidopsis

#### ALIGNMENTS

RESULT 1

Q9TZZ2 PRELIMINARY; PRT; 135 AA.  
AC Q9TZZ2;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE GROUP 2 ALLERGEN EUR M 2 0102 (FRAGMENT).  
GN EUR M 2 0102.  
OS Euroglyphus maynei (House-dust mite).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
OC Acariformes; Sarcotiformes; Astigmata; Analgoidea; Pyroglyphidae;  
OC Euroglyphus.  
OX NCBI\_TaxID=6958;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Smith W., Hart B.J., Thomas W.R.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF047614; AAC82350.1; -  
DR HSSP; P49278; IAGV.  
DR INTERPRO: IPR003172; -  
DR PFAM; PF02221; EL\_Derp2\_DerF2; 1.  
FT NON\_TER  
SQ SEQUENCE 135 AA; 14555 MW; E15E1FA480B8DF16 CRC64;

Query Match 86.1%; Score 68; DB 5; Length 135;  
Best Local Similarity 85.7%; Pred. No. 0.00034;  
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HEIKKLVPGCHGS 14  
|||||:|||||  
Db 17 HEIKKLVPGCHGS 30

RESULT 2

Q96430 PRELIMINARY; PRT; 145 AA.  
ID Q96430  
AC Q96430;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

```

DE  GROUP 2 ALLERGEN EUR M 2 0101.
EUR M 2 0101.
OS  Euroglyphus maynei (House-dust mite).
OC  Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC  Acariformes; Sarcophagales; Astigmata; Analgoidea; Pyroglyphidae;
OC  Euroglyphus.
ON  NCBI_TaxID=6958;
OX  [1]
RN  SEQUENCE FROM N.A.
RP  Smith W., Hart B.J., Thomas W.R.;
RL  Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR  EMBL: AF047613; AAC82349.1; -
DR  HSSP: P49278; 1A9V.
DR  INTERPRO: IPR003172; -
DR  PFAM: PF02221; EL_DerP2_DerF2; 1.
SQ  SEQUENCE 145 AA; 15747 MW; 6655B16C8503A565 CRC64;

Query Match 86.1%; Score 68; DB 5; Length 145;
Best Local Similarity 85.7%; Pred. No. 0.00036;
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HEIKKVLVPGCHGS 14
    |||||:|||| |
DB 27 HEIKKVMVPGCKGS 40

RESULT 3
ID Q9Z090 PRELIMINARY; PRT; 292 AA.
AC Q9Z090;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE TB011.21 PROTEIN.
GN TB011.21.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsi.
OC NCBI_TaxID=3702;
ON [1]
RN SEQUENCE FROM N.A.
RP STRAIN=CV. COLUMBIA;
RC Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.,
RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,
RA Fraser C.M., Venter J.C.;
RT "Arabidopsis thaliana chromosome II BAC TB011 genomic sequence.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC006069; AAD12709.1; -
SQ SEQUENCE 292 AA; 33005 MW; 250E7397BF69569A CRC64;

Query Match 55.7%; Score 44; DB 10; Length 292;
Best Local Similarity 63.6%; Pred. No. 9.6;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 IKKVLVPGCHG 13
    |::|::| |
DB 109 ITKIIVPGSG 119

RESULT 4
ID Q9UQ06 PRELIMINARY; PRT; 1083 AA.
AC Q9UQ06;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE BEC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

QY 1 HEIKVLVPGCHGS 14  
 ID Q9UJG9 PRELIMINARY; PRT; 226 AA.  
 AC Q9UJG9  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE D344A20.1 (NOVEL PROTEIN SIMILAR TO MTHFD1 (METHYLENETETRAHYDROFOLATE  
 DE DEHYDROGENASE (NADP+ DEPENDENT), METHENYLTETRAHYDROFOLATE  
 DE CYCLOHYDROLASE, FORMYLTETRAHYDROFOLATE SYNTHETASE (MTHFC, MTHFD)))  
 DE (FRAGMENT).  
 GN D344A20.1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Matthews L.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL035086; CAB54804.1; -  
 DR HSSP; P11586; 1A41.  
 DR INTERPRO; IPR000672; -  
 DR PFAM; PF00763; THF\_DHG\_CYH; 1.  
 DR PRINTS; PR00085; THFDHDRGNASE.  
 DR KW Hydrolase.  
 FT NON\_TER  
 SQ SEQUENCE 226 AA; 24738 MW; 59844CC5DE5DD723 CRC64;

Query Match 54.4%; Score 43; DB 4; Length 226;  
 Best Local Similarity 72.7%; Pred. No. 11;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 KKVLVPGCHGS 14  
 ID Q45419 PRELIMINARY; PRT; 382 AA.  
 AC Q45419  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE AT2G42640 PROTEIN.  
 GN AT2G42640.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome II of Arabidopsis thaliana."  
 RL Nature 402:761-768(1999).  
 DR EMBL; AC007087; AAD22990.1; -  
 SQ SEQUENCE 382 AA; 41774 MW; 68468D3F6191F120 CRC64;

Query Match 54.4%; Score 43; DB 10; Length 382;  
 Best Local Similarity 53.8%; Pred. No. 19;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 EIKVLVPGCHGS 14  
 ID Q9UFU8 PRELIMINARY; PRT; 917 AA.  
 AC Q9UFU8  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE HYPOTHETICAL 99.4 KDA PROTEIN (FRAGMENT).  
 GN DKFZP586G1517.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Koehler K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL117452; CAB55934.1; -  
 DR HSSP; P11586; 1A41.  
 DR INTERPRO; IPR000559; -  
 DR INTERPRO; IPR000672; -  
 DR PFAM; PF00763; THF\_DHG\_CYH; 1.  
 DR PFAM; PF01268; FTHFS; 1.  
 DR PRINTS; PR00085; THFDHDRGNASE.  
 DR PROSITE; PS00721; FTHFS\_1; 1.  
 DR PROSITE; PS00722; FTHFS\_2; 1.  
 KW Hypothetical protein.  
 FT NON\_TER  
 SQ SEQUENCE 917 AA; 99438 MW; 6AF23EFC2CF02911 CRC64;

Query Match 54.4%; Score 43; DB 4; Length 917;  
 Best Local Similarity 72.7%; Pred. No. 44;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 KKVLVPGCHGS 14  
 ID Q45419 PRELIMINARY; PRT; 697 AA.  
 AC Q45419  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE PUTATIVE TRANSCRIPTIONAL REGULATOR.  
 GN MTLR.  
 OS Bacillus stearothermophilus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1422;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 7954;  
 RX MEDLINE=96421984; PubMed=8824601;  
 RA Henstra S.A., Tolner B., ten Hoeve Duurkens R.H., Konings W.N.,  
 RA Robillard G.T.;  
 RT "Cloning, expression, and isolation of the mannitol transport protein  
 from the thermophilic bacterium Bacillus stearothermophilus.";

RL J. Bacteriol. 178:5586-5591(1996).  
 DR EMBL; U18943; AAC44464.1; -;  
 DR INTERPRO; IPR001550; -;  
 DR PFAM; PF00874; BglG\_antitermin; 2.  
 DR PRODOM; PD004151; -; 1.  
 SQ SEQUENCE 697 AA; 79242 MW; BB052EC2A5C7F94F CRC64;

Query Match 53.2%; Score 42; DB 2; Length 697;  
 Best Local Similarity 58.3%; Pred. No. 50;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 HEIKKVLVPGCH 12  
 |:::|  
 Db 549 HSIKEVLAEACH 560

RESULT 10  
 Q63020 PRELIMINARY; PRT; 276 AA.

ID Q63020  
 AC Q63020;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 DE ALPHA-1-INHIBITOR III (FRAGMENT).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88153707; PubMed=2831216;  
 RA Braciak T.A., Northmann W., Hudson G.O., Shields B.R., Gehring M.R.,  
 RA Fey G.H.;  
 RT "Sequence and acute phase regulation of rat alpha 1-inhibitor III  
 RT messenger RNA.";  
 RL J. Biol. Chem. 263:3999-4012(1988).  
 DR EMBL; M22360; AAA40633.1; -;  
 DR INTERPRO; IPR002890; -;  
 DR PFAM; PF01835; A2M\_N; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 276 AA; 31294 MW; ED18607D639F8175 CRC64;

Query Match 51.9%; Score 41; DB 11; Length 276;  
 Best Local Similarity 85.7%; Pred. No. 30;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 8 VPGCHGS 14  
 |:::|  
 Db 181 VPGCHGN 187

RESULT 11  
 Q9PNM5 PRELIMINARY; PRT; 288 AA.

ID Q9PNM5  
 AC Q9PNM5;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE HYPOTHETICAL PROTEIN CJ1069.  
 GN CJ1069.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 OC Campylobacter.  
 OX NCBI\_TaxID=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=NCTC 11168;  
 RC MERLINE=20150912; PubMed=10688204;  
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
 RA Basham D., Chillingworth T., Davies R.W., Feltwell T., Holtroyd S.,  
 RA Jagers K., Karleyshew A., Moule S., Pallen M.J., Penn C.W.,

RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,  
 RA Whitehead S., Barrell B.G.;  
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
 RT reveals hypervariable sequences.";  
 RL Nature 403:665-668(2000).  
 DR EMBL; AL139077; CAB73324.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 288 AA; 34645 MW; 21EE9C8CF4DC221E CRC64;

Query Match 51.9%; Score 41; DB 2; Length 288;  
 Best Local Similarity 58.3%; Pred. No. 31;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 EIRKVLVPGCHG 13  
 |:::|  
 Db 169 EINKILTKKCHG 180

RESULT 12

O46038 PRELIMINARY; PRT; 475 AA.  
 ID O46038  
 AC O46038;  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)  
 DE EG:103B4.2 PROTEIN.  
 GN EG:103B4.2.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Murphy L., Harris D., Barrell B.;  
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Benos P.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL009193; CAAL5692.1; -;  
 DR FLYBASE; FBgn0023550; EG:103B4.2.  
 SQ SEQUENCE 475 AA; 53756 MW; E78965C742A89A8B CRC64;

Query Match 51.9%; Score 41; DB 5; Length 475;  
 Best Local Similarity 55.8%; Pred. No. 51;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 KKVLPVPGCH 12  
 |:::|  
 Db 343 KNLMPGCH 351

RESULT 13

Q9W508 PRELIMINARY; PRT; 491 AA.  
 ID Q9W508  
 AC Q9W508;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 DE CG18031 PROTEIN.  
 GN CG18031.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananidis P.G., Scher S.E., Li P.W., Hoskins R.A., Gallie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,  
RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,  
RA Abill J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarly C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stappleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RA "The genome sequence of Drosophila melanogaster."  
RT Science 287:2185-2195(2000).  
RL EMBL: AE003423; AAF45750.1; -.  
DR FLYBASE: FBgn0029604; CG18031.  
SQ SEQUENCE 491 AA; 55580 MW; E3CCB33908AF9191 CRC64;

Query Match 51.9%; Score 41; DB 5; Length 491;  
Best Local Similarity 55.6%; Pred. No. 53;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 KKVLPVPGCH 12  
| : : : : :  
DB 343 KNLMIPGCH 351

RESULT 14  
O58294 ID O58294 PRELIMINARY; PRT; 573 AA.  
AC O58294;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE HYPOTHETICAL 64.3 KDA PROTEIN PH0559.  
GN PH0559.  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
OX NCBI\_TaxID=53953;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OT3;  
RX MEDLINE=98344137; PubMed=9679194;  
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,  
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,  
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,

RA Masuchi Y., Shizuya H., Kikuchi H.;  
RT "Complete sequence and gene organization of the genome of a hyper-  
thermophilic archaeobacterium, Pyrococcus horikoshii Orl3.";  
RL DNA res. 5:55-76(1998).  
DR EMBL: AP000002; BAA29648.1; -.  
DR INTERPRO: IPR002814; -.  
DR PFAM: PF01961; DUF110.1.  
KW Hypothetical protein.  
SQ SEQUENCE 573 AA; 64304 MW; 91C98B6BEF45CC36 CRC64;

Query Match 51.9%; Score 41; DB 1; Length 573;  
Best Local Similarity 58.3%; Pred. No. 61;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 IKKVLPGCHGS 14  
| : : : : :  
DB 24 MKKILPGLAGS 35

RESULT 15  
Q9Y606 ID Q9Y606 PRELIMINARY; PRT; 1631 AA.  
AC Q9Y606;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE WUGSC\_H\_RG015P03.1 PROTEIN (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99063792; PubMed=9847074;  
RA Sulston J.E., Waterston R.;  
RT "Toward a complete human genome sequence."  
RL Genome Res. 8:1097-1108(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Leonard S., Wamsley P., Spalding L.;  
RT "The sequence of Homo sapiens BAC clone RG015P03.";  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Waterston R.H.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Waterston R.H.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Waterston R.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC005048; AAD43183.1; -.  
DR HSSP: P02468; 1TLE.  
DR INTERPRO: IPR000561; -.  
DR INTERPRO: IPR001886; -.  
DR INTERPRO: IPR002049; -.  
DR PFAM: PF00053; laminin\_EGF\_12.  
DR PFAM: PF00055; laminin\_Nterm; 1.  
DR PRINTS: PR00011; EGF\_LAMININ.  
DR PROSITE: PS00022; EGF\_1; UNKNOWN\_9.  
DR PROSITE: PS01186; EGF\_2; 1.  
DR PROSITE: PS01248; LAMININ\_TYPE\_EGF; 10.  
FT NON\_TER 1  
SQ SEQUENCE 1631 AA; 178994 MW; 2ADAF6B28E0D69F CRC64;

Query Match 51.9%; Score 41; DB 4; Length 1631;  
Best Local Similarity 63.6%; Pred. No. 1.7e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 4 KKVLPCHGS 14  
:| | | | |  
Db 1329 RKCRPGCHGS 1339

Search completed: April 14, 2001, 10:13:09  
Job time: 214 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2001, 10:11:44 ; Search time 46.83 Seconds  
(without alignments)  
20.545 Million cell updates/sec

Title: US-09-362-731-2

Perfect score: 79

Sequence: 1 HEIKKVLVPGCHGS 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 58722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_67:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	79	100.0	146	2 A60381	allergen Der p II
2	62	78.5	129	2 A61501	allergen Der f II
3	62	78.5	129	2 JU0394	allergen Der f II
4	62	78.5	138	2 B61241	allergen Der f II
5	62	78.5	138	2 A61241	allergen Der f II
6	43	54.4	917	2 T17244	hypothetical prote
7	42	53.2	424	1 XNBVO	ornithine--oxo-aci
8	41	51.9	288	2 H81309	hypothetical prote
9	41	51.9	397	2 S45900	probable membrane
10	41	51.9	573	2 C71170	hypothetical prote
11	41	51.9	1184	2 H70761	probable dnapolyme
12	40	50.6	205	2 E82092	4-methyl-5(β-hydro
13	40	50.6	257	2 S56815	hypothetical prote
14	40	50.6	260	2 A41123	myogenic transcrip
15	40	50.6	631	2 S67268	hypothetical prote
16	40	50.6	1146	2 S59376	hypothetical prote
17	40	50.6	2499	1 A30788	mannose 6-phosphat
18	39.5	50.0	1087	2 T31100	probable potassium
19	39	49.4	325	2 C84060	L-allo-threonine a
20	39	49.4	440	1 XXRTN	phosphatidylcholin
21	39	49.4	2482	2 I48922	cation-independent
22	39	49.4	2483	1 A49617	insulin-like growt
23	38.5	48.7	347	2 T09068	immunophilin-like
24	38.5	48.7	726	2 A47275	transferrin - cock
25	38	48.1	182	2 E72312	conserved hypothet
26	38	48.1	188	2 PL0153	phosphatidylcholin
27	38	48.1	193	2 D83620	hypothetical prote
28	38	48.1	250	2 F72247	methionine aminope
29	38	48.1	323	2 S27907	ATPase - frog viru

30 38 48.1 342 2 E82955 TonB protein PA553  
31 38 48.1 438 1 XXMSN phosphatidylcholin  
32 38 48.1 440 1 XXHUN phosphatidylcholin  
33 38 48.1 440 2 JCI502 phosphatidylcholin  
34 38 48.1 482 1 B69054 DNA-directed DNA p  
35 38 48.1 566 2 E64928 probable acid--CoA  
36 38 48.1 605 2 G72238 lipopolysaccharide  
37 38 48.1 725 2 A45033 myelin transcripti  
38 38 48.1 803 2 A47723 F-spondin precurs  
39 38 48.1 912 1 RDBHNS nitrate reductase  
40 38 48.1 915 1 RDBHNS nitrate reductase  
41 38 48.1 936 2 T06190 lipoxxygenase (EC 1  
42 38 48.1 1023 2 T31669 neural zinc finger  
43 38 48.1 1032 2 T41424 myelin transcripti  
44 38 48.1 1078 2 T42712 myelin transcripti  
45 38 48.1 1184 2 A55184 fibulin-2 precurs

#### ALIGNMENTS

##### RESULT 1

A60381

allergen Der p II precursor - house-dust mite (Dermatophagoides pteronyssinus)  
C:Species: Dermatophagoides pteronyssinus  
C:Date: 03-Mar-1993 #sequence\_revision 03-Mar-1993 #text\_change 13-Sep-1998  
C:Accession: A60381

R:Chua, K.Y.; Doyle, C.R.; Simpson, R.J.; Turner, K.J.; Stewart, G.A.; Thomas, W.R.  
Int. Arch. Allergy Appl. Immunol. 91, 118-123, 1990

A:Title: Isolation of cDNA coding for the major mite allergen Der p II by IgE plaque  
F:18-146/Product: allergen Der p II #status predicted <SIG>  
A:Reference number: A60381; MUID:90256301

A:Accession: A60381

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-146 <CHU>

C:Superfamily: allergen Der p II

F:1-17/Domain: signal sequence #status predicted <SIG>

F:18-146/Product: allergen Der p II #status predicted <MAT>

Query Match 100.0%; Score 79; DB 2; Length 146;

Best Local Similarity 100.0%; Pred. No. 4.4e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HEIKKVLVPGCHGS 14

Db 28 HEIKKVLVPGCHGS 41

##### RESULT 2

A61501

allergen Der f II - house-dust mite (Dermatophagoides farinae) (fragment)

C:Species: Dermatophagoides farinae

C:Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 13-Sep-1998

C:Accession: A61501

R:Trudinger, M.; Chua, K.Y.; Thomas, W.R.

Clin. Exp. Allergy 21, 33-37, 1991

A:Title: cDNA encoding the major mite allergen Der f II.

A:Reference number: A61501; MUID:91215495

A:Accession: A61501

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-129 <TRU>

C:Superfamily: allergen Der p II

Query Match 78.5%; Score 62; DB 2; Length 129;

Best Local Similarity 78.6%; Pred. No. 0.0029;

Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HEIKKVLVPGCHGS 14

Db 11 NEIKKVMVDGCHGS 24



```

RESULT 3
allergen Der f II (pIL2) - house-dust mite (Dermatophagoides farinae)
C:Species: Dermatophagoides farinae
C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 17-Mar-1999
C:Accession: J00394
R:Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haida, M.; Okudaira, H.
Agric. Biol. Chem. 55, 1233-1238, 1991
A:Title: Cloning and expression of cDNA coding for the major house dust mite allergen Der
A:Reference number: PS0417; MUID:91291341
A:Accession: J00394
A:Molecule type: mRNA
A:Residues: 1-129 <YUU>
C:Superfamily: allergen Der p II

Query Match 78.5%; Score 62; DB 2; Length 129;
Best Local Similarity 78.6%; Pred. No. 0.0029;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HEIKKVLVPGCHGS 14
    :|||||:|||||
Db 11 NEIKKVMVDGCHGS 24

RESULT 4
allergen Der f II precursor - house-dust mite (Dermatophagoides farinae) (fragment)
C:Species: Dermatophagoides farinae
C>Date: 12-May-1994 #sequence_revision 27-Jun-1994 #text_change 13-Sep-1998
C:Accession: B61241; J00395
R:Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haida, M.; Dohi, M.; Okudaira
Int. Arch. Allergy Appl. Immunol. 94, 354-356, 1991
A:Title: Synthesis of biologically active recombinant Der f II.
A:Reference number: A61241; MUID:92040281
A:Accession: B61241
A:Molecule type: mRNA
A:Residues: 1-138 <YUU>
C:Superfamily: allergen Der p II
F:1-9/Domain: signal sequence (fragment) #status predicted <SIG>
F:10-138/Product: allergen Der f II #status predicted <MAT>

Query Match 78.5%; Score 62; DB 2; Length 138;
Best Local Similarity 78.6%; Pred. No. 0.003;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HEIKKVLVPGCHGS 14
    :|||||:|||||
Db 20 NEIKKVMVDGCHGS 33

RESULT 5
allergen Der f II precursor - house-dust mite (Dermatophagoides farinae) (fragment)
C:Species: Dermatophagoides farinae
C>Date: 12-May-1994 #sequence_revision 27-Jun-1994 #text_change 13-Sep-1998
C:Accession: A61241; PS0417
R:Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haida, M.; Dohi, M.; Okudaira
Int. Arch. Allergy Appl. Immunol. 94, 354-356, 1991
A:Title: Synthesis of biologically active recombinant Der f II.
A:Reference number: A61241; MUID:92040281
A:Accession: A61241
A:Molecule type: mRNA
A:Residues: 1-138 <YUU>
A:Note: part of this sequence, including the amino end of the mature protein, was confir
C:Superfamily: allergen Der p II
F:1-9/Domain: signal sequence (fragment) #status predicted <SIG>
F:10-138/Product: allergen Der f II #status experimental <MAT>

```

```

Query Match 78.5%; Score 62; DB 2; Length 138;
Best Local Similarity 78.6%; Pred. No. 0.003;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 HEIKKVLVPGCHGS 14
    :|||||:|||||
Db 20 NEIKKVMVDGCHGS 33

```

```

RESULT 6
TI17244
hypothetical protein DKFZp586G1517.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: TI17244
R:Koehler, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wilmann, S.
submitted to the Protein Sequence Database, September 1999
A:Reference number: Z18722
A:Accession: TI17244
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-917 <KOE>
A:Cross-references: EMBL:AL117452
A:Experimental source: adult uterus; clone DKFZp586G1517
C:Genetics:
A:Note: DKFZp586G1517.1
C:Superfamily: C1-tetrahydrofolate synthase; formate--tetrahydrofolate ligase homolog
F:300-917/Domain: formate--tetrahydrofolate ligase homology <FTL>

```

```

Query Match 54.4%; Score 43; DB 2; Length 917;
Best Local Similarity 72.7%; Pred. No. 27;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 4 KKVLPVPGCHGS 14
    :|||:|||||
Db 160 KILVVGCHGS 170

```

```

RESULT 7
XNBYO
ornithine--oxo-acid transaminase (EC 2.6.1.13) - yeast (Saccharomyces cerevisiae)
N:Alternate names: ornithine aminotransferase; ornithine--oxo-acid aminotransferase;
C:Species: Saccharomyces cerevisiae
C>Date: 30-Sep-1991 #sequence_revision 23-Feb-1996 #text_change 21-Jul-2000
C:Accession: S59406; S00181; S05827
R:Du, Z.
submitted to the EMBL Data Library, February 1995
A:Description: The sequence of S. cerevisiae cosmid 9753.
A:Reference number: S59401
A:Accession: S59406
A:Molecule type: DNA
A:Residues: 1-424 <DUZ>
A:Cross-references: EMBL:U01094; NID:g665967; PIDN:AAB67514.1; PID:g665969; GSPDB:GNO
A:Experimental source: strain S288C (AB972)
R:Degols, G.
Eur. J. Biochem. 169, 193-200, 1987
A:Title: Functional analysis of the regulatory region adjacent to the cargB gene of S

```

```

A:Reference number: S00181; MUID:88055042
A:Accession: S00181
A:Molecule type: DNA
A:Residues: 1-6,8-37,'L',39-98,'S',100-211,'R',213-384,'O',386-424 <DEG>
A:Cross-references: EMBL:X06790; NID:g3459; PIDN:CAA29947.1; PID:g3460
R:Degols, G.; Jauniaux, J.C.; Wiame, J.M.
Eur. J. Biochem. 165, 289-296, 1987
A:Title: Molecular characterization of transposable-element-associated mutations that

```

```

A:Reference number: S05827; MUID:87246605
A:Accession: S05827
A:Molecule type: DNA
A:Residues: 1-6,8-37,'L',39-55 <DE2>
A:Cross-references: EMBL:X05571; NID:g3453; PIDN:CAA29081.1; PID:g3454
A:Note: in the authors' translation residues 38-40 are duplicated and, consequently,

```

## C:Genetics:

A:Gene: SGD:CAR2; CARGB; MIPS:YLR438w  
A:Cross-references: SGD:S0004430; MIPS:YLR438w  
A:Map position: 12R  
C:Superfamily: ornithine-oxo-acid aminotransferase  
C:Keywords: aminotransferase; arginine catabolism; phosphoprotein; pyridoxal phosphate  
F:272/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 53.2%; Score 42; DB 1; Length 424;

Best Local Similarity 50.0%; Pred. No. 20;  
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 HEIKKVLVPGCHGS 14

||| |||

Db 288 HDIMSCFTPGSHGS 301

## RESULT 8

H81309  
hypothetical protein Cj1069 [imported] - Campylobacter jejuni (strain NCTC 11168)

C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000

C:Accession: H81309

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barre  
Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp

A:Reference number: A81250; MUID:20150912

A:Accession: H81309

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-288 <PAR>

A:Cross-references: GB:AL139077; GB:AL111168; NID:g6968444; PIDN:CAB73324.1; PID:g696850

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: Cj1069

C:Superfamily: conserved hypothetical protein H10461

Query Match 51.9%; Score 41; DB 2; Length 288;

Best Local Similarity 58.3%; Pred. No. 21;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 EIKKVLVPGCHG 13

||| |||

Db 169 EINKILTKCHG 180

## RESULT 9

S45900

probable membrane protein YBR042c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YBR0412

C:Species: Saccharomyces cerevisiae

C:Date: 26-Aug-1994 #sequence\_revision 09-Sep-1994 #text\_change 22-Oct-1999

C:Accession: S45900

R:Andre, B.; Cziepluch, C.; Hein, C.; Jauniaux, J.C.; Urrestarazu, A.; Vissers, S.

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45893

A:Accession: S45900

A:Molecule type: DNA

A:Residues: 1-397 <AND>

A:Cross-references: EMBL:235911; NID:g536265; PIDN:CAA84984.1; PID:g536266; GSPDB:GN0000

A:Experimental source: strain S288C

C:Genetics:

A:Gene: MIPS:YBR042c

A:Map position: 2R

C:Superfamily: probable membrane protein YBR042c

C:Keywords: transmembrane protein

F:12-37/Domain: transmembrane #status predicted <TM1>

F:55-77/Domain: transmembrane #status predicted <TM2>

F:134-150/Domain: transmembrane #status predicted <TM3>

F:372-390/Domain: transmembrane #status predicted <TM4>

Query Match

Best Local Similarity 51.9%; Score 41; DB 2; Length 397;

Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 HEIKKVLVPG 10

||| |||

Db 8 HKVRKVVP 17

## RESULT 10

C71170

hypothetical protein PH0559 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 21-Jul-2000

C:Accession: C71170

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu  
DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil

A:Reference number: A71000; MUID:98344137

A:Accession: C71170

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-573 <KAW>

A:Cross-references: GB:AP000002; NID:g3236129; PIDN:BAA29648.1; PID:g3256965

A:Experimental source: strain OF3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBa

C:Genetics:

A:Gene: PH0559

Query Match

Best Local Similarity 51.9%; Score 41; DB 2; Length 573;

Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 IKKVLVPGCHGS 14

||| |||

Db 24 MKKILPGLAGS 35

## RESULT 11

H70761

probable dnapolymrase III - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: H70761

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

: Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Sqaes, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987

A:Accession: H70761

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1184 <COL>

A:Cross-references: GB:Z74020; GB:AL123456; NID:g3261584; PIDN:CAA98315.1; PID:g14034

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: dnaE1

C:Superfamily: DNA-directed DNA polymerase III alpha chain

Query Match

Best Local Similarity 51.9%; Score 41; DB 2; Length 1184;

Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 EIKKVLVPGCHGS 14

||| |||

Db 1172 DLKELLPGCLGS 1184

## RESULT 12

E82092  
4-methyl-5(8-hydroxyethyl)-thiazole monophosphate biosynthesis enzyme VC2308 [imported]  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 15-Sep-2000  
C:Accession: E82092  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, H.  
I.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833  
A:Accession: E82092  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-205 <HEI>  
A:Cross-references: GB:AE003852; NID:g9656865; PIDN:AAF95452.1; GSPDB:GN001  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC2308  
A:Map position: 1  
C:Superfamily: signal transduction protein DJ-1

Query Match 50.6%; Score 40; DB 2; Length 205;  
Best Local Similarity 63.6%; Pred. No. 22;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 KKVLPVPGCHG 14  
||| ||| |||  
DB 7 KRILVPVARGH 17

## RESULT 13

S56815  
Hypothetical protein YJL043w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein J1204  
C:Species: Saccharomyces cerevisiae  
C:Date: 08-Jul-1995 #sequence\_revision 08-Sep-1995 #text\_change 05-Nov-1999  
C:Accession: S56815  
R:Pohl, T.M.; Aljinovic, G.  
submitted to the Protein Sequence Database, September 1995  
A:Reference number: S56793  
A:Accession: S56815  
A:Molecule type: DNA  
A:Residues: 1-257 <TOV>  
A:Cross-references: EMBL:Z49318; NID:g1008170; PIDN:CAA89334.1; PID:g1008171; GSPDB:GN00  
C:Genetics:  
A:Gene: YJL043w  
A:Map position: 10L

Query Match 50.6%; Score 40; DB 2; Length 257;  
Best Local Similarity 72.7%; Pred. No. 27;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 EIKKVLVPGCH 12  
|| | ||| ||  
DB 39 EILKSLVPRKH 49

## RESULT 14

A41123  
myogenic transcription factor SUM-1 - sea urchin (Lytechinus variegatus)  
C:Species: Lytechinus variegatus (variegated urchin)  
C:Date: 27-Mar-1992 #sequence\_revision 27-Mar-1992 #text\_change 16-Feb-1997  
C:Accession: A41123  
R:Venutk, J.M.; Goldberg, L.; Chakraborty, T.; Olson, E.N.; Klein, W.H.  
Proc. Natl. Acad. Sci. U.S.A. 86, 6219-6223, 1991  
A:Title: A myogenic factor from sea urchin embryos capable of programming muscle differ  
A:Reference number: A41123; MUID:91296792

A:Accession: A41123  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-260 <VEN>  
A:Cross-references: GB:M69052  
C:Keywords: DNA binding; transcription regulation

Query Match 50.6%; Score 40; DB 2; Length 260;  
Best Local Similarity 58.3%; Pred. No. 28;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 EIKKVLVPGCHG 13  
||: || || ||  
DB 79 ELEHVLAPGFHG 90

## RESULT 15

S67268  
Hypothetical protein YOR356w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein O6629  
C:Species: Saccharomyces cerevisiae  
C:Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 05-Nov-1999  
C:Accession: S67268  
R:Deilus, H.; Hebling, U.; Hofmann, B.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67261  
A:Accession: S67268  
A:Molecule type: DNA  
A:Residues: 1-631 <DEL>  
A:Cross-references: EMBL:Z75264; NID:g1420773; PID:e252180; PID:g1420774; GSPDB:GN000  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: MIPS:YOR356w  
A:Map position: 15R

Query Match 50.6%; Score 40; DB 2; Length 631;  
Best Local Similarity 77.8%; Pred. No. 62;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 VLVPGCHGS 14  
|| || |||||  
DB 242 VLAEGCHGS 250

Search completed: April 14, 2001, 10:11:46  
Job time: 132 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2001, 10:09:34 ; Search time 46.83 Seconds  
(without alignments)  
45.492 Million cell updates/sec

Title: US-09-362-731-1  
Perfect score: 165  
Sequence: 1 QYIKANSKFGITELGSGHEIKKVLVPGCHGS 31  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_67:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84	50.9	146	A60381	allergen Der p II
2	75	45.5	1315	BTCLTN	tentoxylisin (EC 3
3	67	40.6	129	A61501	allergen Der f II
4	67	40.6	129	JU0394	allergen Der f II
5	67	40.6	138	B61241	allergen Der f II
6	67	40.6	138	A61241	allergen Der f II
7	55	33.3	487	S55194	DNA-directed DNA p
8	54	32.7	1333	S38635	blastopia polyprot
9	52	31.5	1386	S73401	MG064 homolog R02
10	51	30.9	332	F69508	N-acetyl-gamma-glu
11	50.5	30.6	244	S29982	class II histocomp
12	50.5	30.6	433	B82537	phosphodiesterase-
13	49	29.7	202	E64362	hypothetical prote
14	49	29.7	267	A64200	DNA polymerase III
15	49	29.7	604	E57372	oligoendopeptidase
16	49	29.7	1882	S73484	hypothetical prote
17	48.5	29.4	237	G75476	conserved hypothet
18	48	29.1	383	C69442	succinyl-CoA synth
19	48	29.1	440	XXRTN	phosphatidylcholin
20	48	29.1	4450	JX0340	gramicidin S synth
21	48	29.1	4452	T49957	gramicidin S synth
22	47	28.5	162	T49957	ribosomal protein
23	47	28.5	188	PL0153	phosphatidylcholin
24	47	28.5	392	B69321	cell division prot
25	47	28.5	438	XXMSN	phosphatidylcholin
26	47	28.5	440	XXHUN	phosphatidylcholin
27	47	28.5	440	JC1502	phosphatidylcholin
28	47	28.5	458	T32634	hypothetical prote
29	47	28.5	6486	T31076	tyrocidine synthet

## ALIGNMENTS

## RESULT 1

A60381  
allergen Der p II precursor - house-dust mite (Dermatophagoides pteronyssinus)  
C;Species: Dermatophagoides pteronyssinus  
C;Date: 03-Mar-1993 #sequence\_revision 03-Mar-1993 #text\_change 13-Sep-1998  
C;Accession: A60381  
R;Chua, K.Y.; Doyle, C.R.; Simpson, R.J.; Turner, K.J.; Stewart, G.A.; Thomas, W.R.  
Int. Arch. Allergy Appl. Immunol. 91, 118-123, 1990  
A;Title: Isolation of cDNA coding for the major mite allergen Der p II by IGE plaque  
A;Reference number: A60381; MUID:90256301  
A;Accession: A60381  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-146 <CHU>  
C;Superfamily: allergen Der p II  
F;1-17/Domain: signal sequence #status predicted <SIG>  
F;18-146/Product: allergen Der p II #status predicted <MAT>

Query Match 50.9%; Score 84; DB 2; Length 146;  
Best Local Similarity 63.6%; Pred. No. 8.7e-05;  
Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 10 IGITELGSGHEIKKVLVPGCHGS 31  
: : : |||||  
Db 20 VDKDCANHEIKKVLVPGCHGS 41

## RESULT 2

BTCLTN  
tentoxylisin (EC 3.4.24.68) precursor - Clostridium tetani  
N;Alternate names: tetanus neurotoxin  
C;Species: Clostridium tetani  
C;Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 18-Jun-1999  
C;Accession: A25689; A25757; A25194; A60759; S69348; S09364  
R;Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, EMBO J. 5, 2495-2502, 1986  
A;Title: Tetanus toxin: primary structure, expression in E. coli, and homology with b  
A;Reference number: A25689; MUID:87053814  
A;Accession: A25689  
A;Molecule type: DNA  
A;Residues: 1-1315 <EIS>  
A;Cross-references: GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770  
R;Fairweather, N.F.; Lyness, V.A.  
Nucleic Acids Res. 14, 7809-7812, 1986  
A;Title: The complete nucleotide sequence of tetanus toxin.  
A;Reference number: A25757; MUID:87040747  
A;Accession: A25757  
A;Molecule type: DNA  
A;Residues: 1-1315 <FAI>  
A;Cross-references: GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774  
A;Experimental source: strain CN3911

class II histocomp  
major merozoite su  
major merozoite su  
major merozoite su  
major merozoite su  
major merozoite su  
major merozoite su  
regulator protein  
ornithine--oxo-aci  
dep protein precu  
bifunctional foli  
transcription init  
vibriolysin (EC 3.  
translation elonga  
hypothetical prote  
H+-transporting AT

30 46.5 28.2 245 2 S29980  
31 46.5 28.2 1060 2 S06286  
32 46.5 28.2 1086 2 S16752  
33 46.5 28.2 1701 2 A54498  
34 46.5 28.2 1701 2 A26868  
35 46.5 28.2 1726 1 S4ZQGM  
36 46.5 28.2 1726 2 A45948  
37 46 27.9 337 2 S44187  
38 46 27.9 424 1 XNBYO  
39 46 27.9 466 2 S36209  
40 46 27.9 554 2 T08976  
41 46 27.9 592 2 S42220  
42 46 27.9 609 1 JT0903  
43 46 27.9 692 2 C64669  
44 46 27.9 1206 2 A64207  
45 45.5 27.6 132 1 PWPFFEL

R;Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.  
 J. Bacteriol. 165, 21-27, 1986  
 A;Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in E  
 A;Reference number: A25194; MUID:86085672  
 A;Accession: A25194  
 A;Molecule type: DNA  
 A;Residues: 743-1315 <FA2>  
 A;Cross-references: GB:M12739; NID:g144920; PIDN:AAA23282.1; PID:g144921  
 A;Accession: B25194  
 A;Molecule type: protein  
 A;Residues: 865-894 <FA3>  
 R;Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.  
 Infect. Immun. 57, 3588-3593, 1989  
 A;Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal  
 A;Reference number: A60759; MUID:90035436  
 A;Accession: A60759  
 A;Molecule type: protein  
 A;Residues: 461-475 <MAT>  
 R;Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.  
 J. Immunol. 142, 394-402, 1989  
 A;Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.  
 A;Reference number: JS0098; MUID:89093918  
 A;Contents: annotation; epitope region  
 R;Schlavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R.  
 Nature 359, 832-835, 1992  
 A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt  
 A;Reference number: S27125; MUID:93063293  
 A;Contents: annotation  
 R;de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.  
 Eur. J. Biochem. 229, 61-69, 1995  
 A;Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.  
 A;Reference number: S69348; MUID:95262688  
 A;Accession: S69348  
 A;Molecule type: protein  
 A;Residues: 2-31 <DEF>  
 C;Comment: The source of this protein was an extrachromosomal plasmid.  
 C;Comment: The precursor is cleaved by endogenous proteinase activity to form light (fra  
 dual chains are not toxic when separated). The amino end of the heavy chain (fragment B)  
 C;Comment: Fragment B forms ion channels in a lipid bilayer. Fragment C binds to ganglio  
 C;Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internalized  
 presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of sy  
 C;Function:  
 A;Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synapt  
 C;Superfamily: tetanus toxin  
 C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc  
 F;2-457/Product: tetroxylisin light chain (fragment A) #status predicted <TTL>  
 F;461-1315/Product: tetroxylisin heavy chain (fragment B.C) #status experimental <TTH>  
 F;461-864/Domain: channel forming (fragment B) #status predicted <TXB>  
 F;865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>  
 F;233,237/Binding site: zinc (His) #status predicted  
 F;234/Active site: Glu #status predicted

Query Match 45.58; Score 75; DB 1; Length 1315;  
 Best Local Similarity 81.08; Pred. No. 0.017;  
 Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 QYIKANSKFIGITELGGHEIK 21  
 |||||  
 Db 830 QYIKANSKFIGITELKKLESK 850

RESULT 3  
 A61501  
 allergen Der f II - house-dust mite (Dermatophagoides farinae) (fragment)  
 C;Species: Dermatophagoides farinae  
 C;Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 13-Sep-1998  
 C;Accession: A61501  
 R;Trudinger, M.; Chua, K.Y.; Thomas, W.R.  
 Clin. Exp. Allergy 21, 33-37, 1991  
 A;Title: cDNA encoding the major mite allergen Der f II.  
 A;Reference number: A61501; MUID:91215495  
 A;Accession: A61501

A;Status: preliminary; not compared with conceptual translation  
 A;Molecule type: mRNA  
 A;Residues: 1-129 <TRU>  
 C;Superfamily: allergen Der p II

Query Match 40.6%; Score 67; DB 2; Length 129;  
 Best Local Similarity 50.0%; Pred. No. 0.021;  
 Matches 11; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Oy 10 IGITELGGHEIKKVLVPGCHGS 31  
 : : : :  
 Db 3 VDVKDCANNEIKKVMVDGCHGS 24

## RESULT 4

JU0394  
 allergen Der f II (pFL2) - house-dust mite (Dermatophagoides farinae)  
 C;Species: Dermatophagoides farinae  
 C;Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 17-Mar-1999  
 C;Accession: JU0394  
 R;Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haida, M.; Okudaira, H.  
 Agric. Biol. Chem. 55, 1233-1238, 1991  
 A;Title: Cloning and expression of cDNA coding for the major house dust mite allergen;  
 A;Reference number: PS0417; MUID:91291341  
 A;Accession: JU0394  
 A;Molecule type: mRNA  
 A;Residues: 1-129 <YUO>  
 C;Superfamily: allergen Der p II

Query Match 40.6%; Score 67; DB 2; Length 129;  
 Best Local Similarity 50.0%; Pred. No. 0.021;  
 Matches 11; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Oy 10 IGITELGGHEIKKVLVPGCHGS 31  
 : : : :  
 Db 3 VDVKDCANNEIKKVMVDGCHGS 24

## RESULT 5

B61241  
 allergen Der f II precursor - house-dust mite (Dermatophagoides farinae) (fragment)  
 C;Species: Dermatophagoides farinae  
 C;Date: 12-May-1994 #sequence\_revision 27-Jun-1994 #text\_change 13-Sep-1998  
 C;Accession: B61241; JU0395  
 R;Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haida, M.; Dohi, M.; Okud  
 Int. Arch. Allergy Appl. Immunol. 94, 354-356, 1991  
 A;Title: Synthesis of biologically active recombinant Der f II.  
 A;Reference number: A61241; MUID:92040281  
 A;Accession: B61241  
 A;Molecule type: mRNA  
 A;Residues: 1-138 <YUO>  
 C;Superfamily: allergen Der p II  
 F;1-9/Domain: signal sequence (fragment) #status predicted <SIG>  
 F;10-138/Product: allergen Der f II #status predicted <MAT>

Query Match 40.6%; Score 67; DB 2; Length 138;  
 Best Local Similarity 50.0%; Pred. No. 0.023;  
 Matches 11; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Oy 10 IGITELGGHEIKKVLVPGCHGS 31  
 : : : :  
 Db 12 VDVKDCANNEIKKVMVDGCHGS 33

## RESULT 6

A61241  
 allergen Der f II precursor - house-dust mite (Dermatophagoides farinae) (fragment)  
 C;Species: Dermatophagoides farinae  
 C;Date: 12-May-1994 #sequence\_revision 27-Jun-1994 #text\_change 13-Sep-1998  
 C;Accession: A61241; PS0417

R.Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haida, M.; Dohi, M.; Okudaira Int. Arch. Allergy Appl. Immunol. 94, 354-356, 1991  
 A:Title: Synthesis of biologically active recombinant Der f II.  
 A:Reference number: A61241; MUID:92040281  
 A:Accession: A61241  
 A:Molecule type: mRNA  
 A:Residues: 1-138 <YU>  
 A:Note: part of this sequence, including the amino end of the mature protein, was confirmed by NMR.  
 C:Superfamily: allergen Der p II  
 F:1-9/Domain: signal sequence (fragment) #status predicted <SIG>  
 F:10-138/Product: allergen Der f II #status experimental <MAT>

Query Match 40.6%; Score 67; DB 2; Length 138;  
 Best Local Similarity 50.0%; Pred. No. 0.023;  
 Matches 11; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 10 IGITELGGHEIKKVLVPGCHGS 31  
 Db 12 VDVKDCANNEIKKVMVDGCHGS 33

RESULT 7  
 S55194  
 DNA-directed DNA polymerase (EC 2.7.7.7) III regulatory chain - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: DNA-directed DNA polymerase delta small chain; HUS2 protein; HYS2 protein  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 30-Oct-1998 #sequence\_revision 30-Oct-1998 #text\_change 21-Jul-2000  
 C:Accession: S55194; S57021; S59122  
 R:de Haan, M.; Smits, P.H.M.; Grivell, L.A.  
 submitted to the EMBL Data Library, May 1995  
 A:Reference number: S55183  
 A:Accession: S55194  
 A:Molecule type: DNA  
 A:Residues: 1-487 <DE>  
 A:Cross-references: EMBL:X87611; NID:g854567; PIDN:CAA60928.1; PID:g854579  
 R:de Haan, M.; Grivell, L.A.; Smits, P.H.M.  
 submitted to the Protein Sequence Database, September 1995  
 A:Reference number: S56771  
 A:Accession: S57021  
 A:Molecule type: DNA  
 A:Residues: 1-487 <ZAG>  
 A:Cross-references: EMBL:Z49506; NID:g1015629; PIDN:CAA89528.1; PID:g1015630; GSPDB:GN00  
 R:Sugimoto, K.; Sakamoto, Y.; Takahashi, O.; Matsumoto, K.  
 Nucleic Acids Res. 23, 3493-3500, 1995  
 A:Title: HYS2, an essential gene required for DNA replication in Saccharomyces cerevisiae  
 A:Reference number: S59122; MUID:96032843  
 A:Accession: S59122  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-155, 'H', 157-464, 'N', 466-487 <SUG>  
 A:Cross-references: EMBL:D50324; NID:g987711; PIDN:BAAO8859.1; PID:g1256943  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995  
 C:Genetics:  
 A:Gene: SGD:HUS2; HYS2; MIPS:YJR006w  
 A:Cross-references: SGD:S0003766; MIPS:YJR006w  
 A:Map position: 10R  
 C:Complex: heterodimer of catalytic (see PIR:RNBVL3) and regulatory chain  
 C:Function:  
 A:Description: DNA polymerase delta is involved in DNA replication and in nucleotide excision repair  
 C:Superfamily: human DNA-directed DNA polymerase delta regulatory chain  
 C:Keywords: DNA repair; DNA replication; nucleotidyltransferase; nucleus

Query Match 33.3%; Score 55; DB 1; Length 487;  
 Best Local Similarity 44.4%; Pred. No. 4.5;  
 Matches 12; Conservative 4; Mismatches 9; Indels 2; Gaps 1;

QY 2 YIKANSKFTG--ITELGGHEIKKVLVLP 26  
 Db 436 YIVANQPIYFGRVVEIGKNNIISVP 462

## RESULT 8

S38635  
 blastopodia polypeptide - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 17-Nov-2000  
 C:Accession: S38635  
 R:Frommer, G.; Schuh, R.; Jdckle, H.  
 submitted to the EMBL Data Library, November 1993  
 A:Description: Localized expression of a novel microRNA-like element in the blastoder  
 A:Reference number: S38635  
 A:Accession: S38635  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1333 <PRO>  
 A:Cross-references: EMBL:Z27119; NID:g415797; PID:g415798  
 C:Genetics:  
 A:Gene: FlyBase:micropia  
 A:Cross-references: FlyBase:FBgn0014947  
 C:Keywords: polypeptide

Query Match 32.7%; Score 54; DB 2; Length 1333;

Best Local Similarity 35.9%; Pred. No. 18;

Matches 14; Conservative 7; Mismatches 10; Indels 8; Gaps 1;

QY 1 QYIKANSKFTGTEL-----GGHEIKKVLVPGCHGS 31

Db 127 KYQVARSKMIGSAELFLESECVSGYTELKELLIEEFGS 165

## RESULT 9

S73401  
 MG064 homolog R02.orf1386V - Mycoplasma pneumoniae (strain ATCC 29342)  
 C:Species: Mycoplasma pneumoniae  
 A:Variety: ATCC 29342  
 C:Date: 26-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 07-Dec-1999  
 C:Accession: S73401  
 R:Himmelsreith, R.; Hilbert, H.; Piagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.  
 Nucleic Acids Res. 24, 4420-4449, 1996  
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae  
 A:Reference number: S73327; MUID:97105885  
 A:Accession: S73401  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1386 <HIM>  
 A:Cross-references: EMBL:AE000009; GB:U00089; NID:g1673720; PIDN:AAB95723.1; PID:g167  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996  
 C:Genetics:  
 A:Genetic code: SGC3  
 A:Start codon: GTG

Query Match 31.5%; Score 52; DB 2; Length 1386;

Best Local Similarity 34.2%; Pred. No. 36;

Matches 13; Conservative 7; Mismatches 10; Indels 8; Gaps 1;

QY 2 YIKANSKFTGTELGGHEIKKV-----LVPGCHGS 31

Db 538 YLKNTEQIGILKANGVSGRKINLSLIFSLIPGIVGS 575

## RESULT 10

F69508  
 N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) - Archaeoglobus fulgidus  
 C:Species: Archaeoglobus fulgidus  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: F69508  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
 : Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E  
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,  
 Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silva M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.  
M.: Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF2599

Query Match 30.6%; Score 50.5; DB 2; Length 433;  
Best Local Similarity 32.3%; Pred. No. 18;  
Matches 10; Conservative 10; Mismatches 10; Indels 1; Gaps 1;

Qy 2 YKANSPFPGITELGGHEIK-KVLVPGCHGS 31  
:: :: : ||| | ||| | : : | :  
Db 273 HWSDGQVLGITPLPGHESKVEAMLLGAHAT 303

RESULT 13  
E64362  
hypothetical protein MJ0501 - Methanococcus jannaschii  
C:Species: Methanococcus jannaschii  
C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
C:Accession: E64362  
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,  
rson, J.D.; Sadov, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,  
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
A:Reference number: A64300; MUID:96337999  
A:Accession: E64362  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-202 <BUL>  
A:Cross-references: GB:U67500; GB:L77117; NID:g1591202; PIDN:AAB98492.1; PID:g1591200  
C:Genetics:  
A:Map position: REV443001-442393  
A:Start codon: GTG  
C:Superfamily: hypothetical protein MJ0501

Query Match 29.7%; Score 49; DB 2; Length 202;  
Best Local Similarity 37.5%; Pred. No. 13;  
Matches 9; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Qy 8 KFIGITELGGHEIKKVLVPGCHGS 31  
| : | : | : : ||| : | : ||:  
Db 2 KIMEIFPKNGCVKKLFIQGLHGN 25

RESULT 14  
A64200  
DNA polymerase III beta chain - Mycoplasma genitalium  
C:Species: Mycoplasma genitalium  
C>Date: 17-Nov-1995 #sequence\_revision 17-Nov-1995 #text\_change 19-May-2000  
C:Accession: A64200  
R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.  
M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick  
, C.A.; Venter, J.C.  
Science 270, 397-403, 1995  
A:Title: The minimal gene complement of Mycoplasma genitalium.  
A:Cross-references: GB:U39679; GB:I43967; NID:g1045668; PID:g1045669; TIGR:MG001  
A:Reference number: A64200; MUID:96026346  
A:Accession: A64200  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-267 <TIGR>  
A:Cross-references: GB:U39679; GB:I43967; NID:g1045668; PID:g1045669; TIGR:MG001  
A:Experimental source: strain G-37  
C:Genetics:  
A:Gene: dnaN  
A:Genetic code: SGC3  
A:Start codon: GTG  
C:Superfamily: DNA-directed DNA polymerase III beta chain



Query Match 29.7%; Score 49; DB 2; Length 267;  
Best Local Similarity 43.5%; Pred. No. 18;  
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 4 KANSKFIGITELGGHEIKKVLVP 26  
: |||: || |::|||  
Db 243 QGNSKYFLITSKSEPELKQILVP 265

## RESULT 15

E75372  
oligoendopeptidase - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C:Accession: E75372  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: E75372  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-604 <WHI>  
A:Cross-references: GB:AE002007; GB:AE00513; NID:g6459402; PIDN:AAF11188.1; PID:g645939  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR1627  
A:Map position: 1  
C:Superfamily: oligoendopeptidase F

Query Match 29.7%; Score 49; DB 2; Length 604;  
Best Local Similarity 34.6%; Pred. No. 41;  
Matches 9; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

QY 1 QYIKANSKFIGITELGGHEIKKVLVP 26  
: | | ::||:| ::| |||  
Db 282 RYWKVRDNLGLSELREYDVKSLVP 307

Search completed: April 14, 2001, 10:11:44  
Job time: 130 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2001, 10:11:58 ; Search time 27.31 seconds  
(without alignments)  
38.884 Million cell updates/sec

Title: US-09-362-731-1

Perfect score: 165

Sequence: 1 QYIKANSKFIGITELGGHEIKVLVPGCHGS 31

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	84	50.9	146	DER2_DERPT	P49278 dermatophag
2	75	45.5	1314	TETX_CLOTE	P04958 clostridium
3	67	40.6	146	DEF2_DERFA	Q00855 dermatophag
4	55	33.3	487	DPD2_YEAST	P46957 saccharomyc
5	52	31.5	1386	Y064_MYCPN	P75613 mycoplasma
6	51	30.9	332	ARGC_ARCFU	O28208 archaeoglob
7	50.5	30.6	347	ARGC_CORGL	Q59279 corynebacte
8	49	29.7	202	Y501_METJA	Q57924 methanococc
9	49	29.7	259	LPXA_CHRVI	Q46481 chromatium
10	49	29.7	364	DP3B_MYCGE	P47247 mycoplasma
11	49	29.7	1882	Y468_MYCPN	P75109 mycoplasma
12	48.5	29.4	167	Y742_CAEEL	Q11070 caenorhabdi
13	48	29.1	383	SUCL_ARCFU	O28732 archaeoglob
14	48	29.1	440	LCAT_RAT	P18424 rattus norv
15	48	29.1	4451	GRSB_BACBR	P14688 b gramicidi
16	47	28.5	188	LCAT_PIG	P30930 sus scrofa
17	47	28.5	392	FTZ2_ARCFU	O29685 archaeoglob
18	47	28.5	438	LCAT_MOUSE	P16301 mus musculu
19	47	28.5	440	LCAT_HUMAN	P04180 homo sapien
20	47	28.5	440	LCAT_PAPAN	Q08758 papio anubi
21	47	28.5	440	LCAT_RABIT	P35761 oryctolagus
22	47	28.5	6486	TYOC_BACBR	Q30409 b tyrocidin
23	46.5	28.2	207	REQQ_BPPS3	O80286 bacterioph
24	46.5	28.2	1682	MSPL_PLAF3	P19598 plasmodium
25	46.5	28.2	1701	MSPL_PLAFF	P13819 plasmodium
26	46.5	28.2	1701	MSPL_PLAFM	P08569 plasmodium
27	46.5	28.2	1726	MSPL_PLAFC	P04934 plasmodium
28	46.5	28.2	1726	MSPL_PLAFC	P50495 plasmodium
29	46	27.9	337	MAIR_STAXY	Q56201 staphylococ
30	46	27.9	424	OAT_YEAST	P07991 saccharomyc
31	46	27.9	499	SYFA_THEAC	P57693 thermoplas
32	46	27.9	592	T2D5_DROME	P49847 drosophila
33	46	27.9	609	NPRV_VIBPR	Q00971 vibrio prot

34 46 27.9 691 1 EFG\_HELPY  
35 46 27.9 1331 1 Y064\_MYCGE  
36 45.5 27.6 132 1 ATPE\_PYLLI  
37 45.5 27.6 1235 1 TRK1\_YEAST  
38 45.5 27.6 1241 1 TRK1\_SACBA  
39 45 27.3 141 1 ALG2\_TYRPU  
40 45 27.3 287 1 TRUB\_AQUAE  
41 45 27.3 559 1 ENS2\_YEAST  
42 45 27.3 691 1 EFG\_HELPY  
43 45 27.3 872 1 DNAB\_SYNY3  
44 44.5 27.0 155 1 BCP\_HAEIN  
45 44.5 27.0 207 1 REGQ\_BPHK0

P56002 helicobacte  
P47310 mycoplasma  
P26534 pylaiella 1  
P12685 saccharomyc  
P28569 saccharomyc  
O02380 tyrophagus  
O06922 aquifex aeo  
P12294 saccharomyc  
Q95418 synecocyst  
Q54118 helicobacte  
P44411 haemophilus  
Q02582 bacterioph

#### ALIGNMENTS

RESULT 1  
DER2\_DERPT STANDARD; PRT; 146 AA.  
AC P49278;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE MITE ALLERGEN DER P 2 PRECURSOR (DER P II) (DPX).  
GN DERP2.  
OS Dermatophagoides pteronyssinus (House-dust mite).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
OC Acariformes; Sarcotiformes; Astigmata; Analgoidea; Pyroglyphidae;  
OC Dermatophagoides.  
OX NCBI\_TaxID=6956;  
RN [1]  
RX SEQUENCE FROM N.A.  
RX MEDLINE=90256301; PubMed=2341191;  
RA Chua K.Y., Doyle C.R., Simpson R.J., Turner K.J., Stewart G.A.,  
RA Thomas W.R.;  
RT "Isolation of cDNA coding for the major mite allergen Der p II by IgE  
RT plaque immunoassay.";  
RL Int. Arch. Allergy Appl. Immunol. 91:118-123(1990).  
RN [2]  
RP PARTIAL SEQUENCE OF 18-57.  
RX MEDLINE=89278484; PubMed=2732406;  
RA Heymann P.W., Chapman M.D., Aalberse R.C., Fox J.W.,  
RA Platts-Mills T.A.;  
RT "Antigenic and structural analysis of group II allergens (Der f II  
RT and Der p II) from house dust mites (Dermatophagoides spp).";  
RL J. Allergy Clin. Immunol. 83:1055-1067(1989).  
RN [3]  
RP VARIANTS.  
RA Smith W., Hales B.J., Thomas W.R.;  
RT "Allergens of wild house dust mites: environmental Der p 1 and Der p 2  
RT sequence polymorphisms.";  
RL Submitted (JUN-2000) to the SWISS-PROT data bank.  
RN [4]  
RP STRUCTURE BY NMR.  
RX MEDLINE=98409423; PubMed=9737847;  
RA Mueller G.A., Benjamin D.C., Rule G.S.;  
RT "Tertiary structure of the major house dust mite allergen Der p 2:  
RT sequential and structural homologies.";  
RL Biochemistry 37:12707-12714(1998).  
CC -I- SIMILARITY: TO MITE ALLERGEN LEP D I.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; AF276239; AAF86462.1; -  
CC PDB; LA9V; 14-OCT-98.  
CC Allergen; Signal; 3D-structure; Polymorphism.  
KW

FT SIGNAL 1 17  
 FT CHAIN 18 146  
 FT DISULFID 25 136  
 FT DISULFID 38 44  
 FT DISULFID 90 95  
 FT VARIANT 39 39  
 FT VARIANT 40 40  
 FT VARIANT 44 44  
 FT VARIANT 47 47  
 FT VARIANT 49 49  
 FT VARIANT 56 56  
 FT VARIANT 61 61  
 FT VARIANT 75 75  
 FT VARIANT 78 78  
 FT VARIANT 81 81  
 FT VARIANT 95 95  
 FT VARIANT 98 98  
 FT VARIANT 108 108  
 FT VARIANT 111 111  
 FT VARIANT 114 114  
 FT VARIANT 116 116  
 FT VARIANT 118 118  
 FT VARIANT 127 127  
 SQ SEQUENCE 146 AA; 15999 MW; 59182FA7FD26D3AF CRC64;

Query Match 50.9%; Score 84; DB 1; Length 146;  
 Best Local Similarity 63.6%; Pred. No. 1.5e-05;  
 Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 10 IGITELGGHEIKKVLVPGCHGS 31  
 : : : |||||  
 Db 20 VDVKDCANHEIKKVLVPGCHGS 41

RESULT 2  
 TETX\_CLOTE STANDARD; PRT; 1314 AA.  
 AC P04958;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE TETANUS TOXIN PRECURSOR (EC 3.4.24.68) (TENTOXILYSIN).  
 OS Clostridium tetani.  
 OG Plasmid.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1513;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=87053814; PubMed=3536478;  
 RA Elsel U., Jarausch W., Goretzki K., Henschen A., Engels J.,  
 RA Weller U., Hudel M., Habermann E., Niemann H.;  
 RT "Tetanus toxin: primary structure, expression in E. coli, and  
 RT homology with botulinum toxins";  
 RL EMBO J. 5:2495-2502(1986).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=CN3911;  
 RX MEDLINE=87040747; PubMed=3774547;  
 RA Fairweather N.F., Lyness V.A.;  
 RA "The complete nucleotide sequence of tetanus toxin";  
 RL Nucleic Acids Res. 14:7809-7812(1986).  
 RN [3]  
 RN SEQUENCE OF 742-1314 FROM N.A.  
 RX MEDLINE=86085672; PubMed=3510187;  
 RA Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;  
 RT "Cloning, nucleotide sequencing, and expression of tetanus toxin  
 RT fragment C in Escherichia coli";  
 RL J. Bacteriol. 165:21-27(1986).  
 RN [4]  
 RN PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RP PRINTS; PR00760; BONTOXILYSIN.  
 RX MEDLINE=9201034; PubMed=2108021;

RA Kriegstein K., Henschen A., Weller U., Habermann E.;  
 RT "Arrangement of disulfide bridges and positions of sulfhydryl groups  
 RL in tetanus toxin";  
 RN Eur. J. Biochem. 188:39-45(1990).  
 RP [5]  
 RP PARTIAL SEQUENCE.  
 RX MEDLINE=92037649; PubMed=1935979;  
 RA Kriegstein K.G., Henschen A.H., Weller U., Habermann E.;  
 RT "Limited proteolysis of tetanus toxin. Relation to activity and  
 RT identification of cleavage sites";  
 RL Eur. J. Biochem. 202:41-51(1991).  
 RN [6]  
 RP IDENTIFICATION AS ZINC-PROTEASE.  
 RX MEDLINE=93010948; PubMed=1396558;  
 RA Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,  
 RA Montecucco C.;  
 RT "Tetanus toxin is a zinc protein and its inhibition of  
 RT neurotransmitter release and protease activity depend on zinc";  
 RL EMBO J. 11:3577-3583(1992).  
 RN [7]  
 RP IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE=93063293; PubMed=1331807;  
 RA Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,  
 RA Dasgupta B.R., Montecucco C.;  
 RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release  
 RT by proteolytic cleavage of synaptobrevin";  
 RL Nature 359:832-835(1992).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.  
 RX MEDLINE=97475217; PubMed=9334741;  
 RA Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,  
 RA Sax M.;  
 RT "Structure of the receptor binding fragment HC of tetanus  
 RL neurotoxin";  
 RL Nat. Struct. Biol. 4:788-792(1997).  
 CC -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-1-PHE-77  
 CC BOND OF SYNAPTOSOMAL-2.  
 CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-1-PHE-77 BOND IN  
 CC SYNAPTOSOMAL-2.  
 CC -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO  
 CC YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE  
 CC AND ARE NON-TOXIC AFTER SEPARATION.  
 CC -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO  
 CC GANGLIOSIDE RECEPTORS.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC  
 CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X04436; CAA28033.1; -;  
 DR EMBL; M12739; AAA23282.1; -;  
 DR EMBL; X06214; CAA29564.1; -;  
 DR PIR; A25689; BTCLTN.  
 DR PDB; 1AF9; 29-APR-98.  
 DR PDB; 1A8D; 14-OCT-98.  
 DR MEROPS; M27.001; -;  
 DR InterPro; IPR000130; -;  
 DR InterPro; IPR000395; -;  
 DR Pfam; PF01742; Peptidase\_M27; 1.  
 DR PRINTS; PR00760; BONTOXILYSIN.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.

KW Newkotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;  
 FT 3D-structure. 0 0  
 FT INIT\_MET 0 0  
 FT CHAIN 457 1314  
 FT CHAIN 457 1314  
 FT METAL 232 232  
 FT ACT\_SITE 233 233  
 FT METAL 236 236  
 FT TRANSMEM 226 246  
 FT TRANSMEM 669 689  
 FT DISULFID 438 466  
 FT DISULFID 1076 1092  
 SQ SEQUENCE 1314 AA; 150550 MW; 134C3657133EF81D CRC64;  
 Query Match 45.5%; Score 75; DB 1; Length 1314;  
 Best Local Similarity 81.0%; Pred. No. 0.0035;  
 Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 QYIKANSKFGITELGHEIK 21  
 Db 829 QYIKANSKFGITELKLESK 849  
 RESULT 3  
 DEF2\_DERFA STANDARD; PRT; 146 AA.  
 ID DEF2\_DERFA  
 AC Q00855; P39672; Q26359;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DE 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE MITE ALLERGEN DER F 2 PRECURSOR (DER F II).  
 GN DERF2.  
 OS Dermatophagoides farinae (House-dust mite).  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
 OC Acariformes; Sarcotiformes; Astigmata; Analgoidea; Pyroglyphidae;  
 OC Dermatophagoides.  
 OX NCBI\_TaxID=6954;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=91291341; PubMed=1368682;  
 RA Yuuki T., Okumura Y., Ando T., Yamakawa H., Suko M., Haida M.,  
 RA Okudaira H.;  
 RT "Cloning and expression of cDNA coding for the major house dust mite  
 RT allergen Der f II in *Escherichia coli*.";  
 RL Agric. Biol. Chem. 55:1233-1238(1991).  
 RN [2]  
 RP SEQUENCE OF 4-146 FROM N.A.  
 RX MEDLINE=94256850; PubMed=8198452;  
 RA Okuhira H.;  
 RT "Molecular biology of mite antigens.";  
 RL Arerugi 43:435-440(1994).  
 RN [3]  
 RP DISULFIDE BONDS, AND PARTIAL SEQUENCE.  
 RX MEDLINE=93283958; PubMed=8508052;  
 RA Nishiyama C., Yuuki T., Takai T., Okumura Y., Okudaira H.;  
 RT "Determination of three disulfide bonds in a major house dust mite  
 RT allergen, Der f II.";  
 RL Int. Arch. Allergy Immunol. 101:159-166(1993).  
 RN [4]  
 RP PARTIAL SEQUENCE OF 18-52.  
 RX MEDLINE=89278484; PubMed=2732406;  
 RA Heymann P.W., Chapman M.D., Aalberse R.C., Fox J.W.,  
 RA Platts-Mills T.A.;  
 RT "Antigenic and structural analysis of group II allergens (Der f II  
 RT and Der p II) from house dust mites (Dermatophagoides spp).";  
 RL J. Allergy Clin. Immunol. 83:1055-1067(1989).  
 RN [5]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE=98079068; PubMed=9417088;  
 RA Ichikawa S., Hatanaka H., Yuuki T., Iwamoto N., Kojima S.,  
 RA Nishiyama C., Ogura K., Okumura Y., Inagaki F.;  
 RT "Solution structure of Der f 2, the major mite allergen for atopic

RT diseases.";  
 RL J. Biol. Chem. 273:356-360(1998).  
 CC -!- MISCELLANEOUS: THE SEQUENCE SHOWN HERE IS FROM CLONE 2. THE  
 CC N-TERMINAL SEQUENCE (AA 1-8) FROM CLONE 1 AND 11 ARE NOT YET  
 CC KNOWN.  
 CC -!- SIMILARITY: TO MITE ALLERGEN LEP D I.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D10447; BAA01239.1; -  
 DR EMBL; D10448; BAA01240.1; -  
 DR EMBL; D10449; BAA01241.1; -  
 DR EMBL; S70378; AAB30829.1; -  
 DR PIR; PS0417; PS0417.  
 DR PDB; LAHK; 08-APR-98.  
 DR PDB; LAHM; 08-APR-98.  
 KW Allergen; Signal; Polymorphism; 3D-structure.  
 FT SIGNAL 1 17  
 FT CHAIN 18 146 MITE ALLERGEN DER F 2.  
 FT DISULFID 25 136  
 FT DISULFID 38 44  
 FT DISULFID 90 95  
 FT VARIANT 93 93 M -> V (IN CLONE 1).  
 FT VARIANT 105 105 I -> A (IN CLONE 1).  
 FT VARIANT 128 128 I -> V (IN CLONE 1).  
 FT VARIANT 142 142 G -> A (IN CLONE 1).  
 FT CONFLICT 5 8 ILC -> GTMV (IN REF. 2).  
 SQ SEQUENCE 146 AA; 15802 MW; FA118206CD88534A CRC64;  
 Query Match 40.6%; Score 67; DB 1; Length 146;  
 Best Local Similarity 50.0%; Pred. No. 0.0054;  
 Matches 11; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
 QY 10 IGITELGGHEIKKVLVPGCHGS 31  
 Db 20 VDVKDCANNEIKKVMVDGCHGS 41  
 RESULT 4  
 DP2\_YEAST STANDARD; PRT; 487 AA.  
 ID DP2\_YEAST  
 AC P46957;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE DNA POLYMERASE DELTA SMALL SUBUNIT (EC 2.7.7.7).  
 GN POL31 OR HYS2 OR HUS2 OR SDP5 OR YJR006W OR J1427 OR YJR83.7.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sugimoto K., Sakamoto Y., Matsumoto K.;  
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / FY1679;  
 RA de Haan M., Smits P.H.M., Grivell L.A.;  
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION.  
 RX MEDLINE=98083176; PubMed=9421503;  
 RA Hashimoto K., Nakashima N., Ohara T., Maki S., Sugino A.;  
 RT "The second subunit of DNA polymerase III (delta) is encoded by the  
 RT HYS2 gene in *Saccharomyces cerevisiae*.";



DR EMBL; AF000961; AAB89185.1; -  
 DR TIGR; AF2071; -  
 DR Pfam; PF01118; Semialdehyde\_dh; 1.  
 DR PROSITE; PS01224; ARG; 1.  
 KW Arginine biosynthesis; Oxidoreductase; NADP.  
 FT ACT\_SITE 144 144 BY SIMILARITY  
 SQ SEQUENCE 332 AA; 37304 MW; 295EB7504EECE33 CRC64;

Query Match 30.9%; Score 51; DB 1; Length 332;  
 Best Local Similarity 45.2%; Pred. No. 3.3;  
 Matches 14; Conservative 4; Mismatches 7; Indels 7; Gaps 2;

QY 2 YKANSKFIGITELGHEIKK---VLVPGCH 29  
 ||| |::| |::| |::| |::|  
 Db 118 YVEA---VYGLTELHREIKKANLVANPGCY 145

RESULT 7  
 ARGCC\_CORGL STANDARD; PRT; 347 AA.  
 ID ARGCC\_CORGL  
 AC Q59279; O32353;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE N-ACETYL-GAMMA-GLUTAMYL-PHOSPHATE REDUCTASE (EC 1.2.1.38) (AGPR) (N-  
 ACETYL-GLUTAMATE SEMIALDEHYDE DEHYDROGENASE) (NAGSA DEHYDROGENASE).  
 GN ARGCC.

OS Corynebacterium glutamicum (Brevibacterium flavum).  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
 OC Corynebacterium.  
 OX NCBI\_TaxID=1718;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 13059 / AS019;  
 RA Chun J.Y., Lee E.J., Cheon C.I., Min K.H., Lee M.S.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 13059 / AS019;  
 RA Park M.Y., Chun J.Y., Ko S.Y., Lee M.S.;  
 RT "Molecular cloning of the arginine biosynthetic genes from  
 Corynebacterium glutamicum".  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]

RP SEQUENCE OF 243-347 FROM N.A.  
 RC STRAIN-ATCC 13032;  
 RX MEDLINE=96146054; PubMed=8581175;  
 RA Sakanyan V., Petrosyan P., Lecoq M., Boyen A., Legrain C.,  
 RA Demarez M.N., Hallet J.N., Glansdorff N.;  
 RT "Genes and enzymes of the acetyl cycle of arginine biosynthesis in  
 Corynebacterium glutamicum: enzyme evolution in the early steps of  
 the arginine pathway."  
 RT Microbiology 142:99-108(1996).  
 RL CATALYTIC ACTIVITY: N-ACETYL-L-GLUTAMATE 5-SEMIALDEHYDE + NADP(+)  
 + ORTHOPHOSPHATE = N-ACETYL-5-GLUTAMYL PHOSPHATE + NADPH.  
 CC -!- PATHWAY: THIRD STEP IN ARGININE BIOSYNTHESIS.  
 CC -!- SIMILARITY: BELONGS TO THE NAGSA DEHYDROGENASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF005242; AAB62245.1; -  
 DR EMBL; AF049897; AAC24812.1; -  
 DR EMBL; X86157; CAA60096.1; -  
 DR InterPro; IPR000706; -  
 DR Pfam; PF01118; Semialdehyde\_dh; 1.

DR PROSITE; PS01224; ARG; 1.  
 KW Arginine biosynthesis; Oxidoreductase; NADP.  
 FT ACT\_SITE 151 151 BY SIMILARITY.  
 FT CONFLICT 277 277 S -> L (IN REF. 3).  
 FT CONFLICT 341 341 A -> P (IN REF. 3).  
 SQ SEQUENCE 347 AA; 35910 MW; 98D333C1CE8A4D98F CRC64;

Query Match 30.6%; Score 50.5; DB 1; Length 347;  
 Best Local Similarity 47.8%; Pred. No. 4.1;  
 Matches 11; Conservative 2; Mismatches 5; Indels 5; Gaps 1;

QY 11 GITELGGHE-----IKKVLVPGC 28  
 ||| |::| |::| |::| |::|  
 Db 129 GIPMPGHREALGAKRVAVPGC 151

RESULT 8  
 Y501\_METJA STANDARD; PRT; 202 AA.  
 ID Y501\_METJA  
 AC Q57924;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE HYPOTHETICAL PROTEIN MJ0501.  
 GN MJ0501.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
 OC Methanococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kervilave A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.P., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii."  
 RL Science 273:1058-1073(1996).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U67500; AAB98492.1; -  
 DR TIGR; MJ0501; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 202 AA; 23743 MW; 8F407212641089A5 CRC64;

Query Match 29.7%; Score 49; DB 1; Length 202;  
 Best Local Similarity 37.5%; Pred. No. 3.8;  
 Matches 9; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 8 KFIGITELGGHEIKKVLVPGCHGS 31  
 |::| |::| |::| |::|  
 Db 2 KIMEIFEFGKGVKRLFTIGLHGN 25

RESULT 9  
 LPXA\_CHRVI STANDARD; PRT; 259 AA.  
 ID LPXA\_CHRVI  
 AC Q46481;









```

CC *INCORPORATES ONE AMINO ACID INTO THE PEPTIDE PRODUCT AND CAN BE
CC FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR SUBSTRATE
CC ADENYLATION, THIOLATION, CONDENSATION (NOT FOR THE INITIATION
CC MODULE), AND EPIMERIZATION (OPTIONAL), AND N METHYLATION
CC (OPTIONAL).
CC -!- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X61658; CAA43838.1; -
DR EMBL; M29703; AAA58719.1; -
DR EMBL; X15577; CAA33604.1; -
DR EMBL; D00938; BAA00778.1; -
DR PIR; S20542; YGBSG2.
DR HSP; P14687; IAMU.
DR InterPro; IPR000255; -
DR InterPro; IPR000873; -
DR InterPro; IPR001031; -
DR InterPro; IPR001242; -
DR Pfam; PF00501; AMP-binding; 4.
DR Pfam; PF00668; DUF4; 4.
DR Pfam; PF00975; Thioesterase; 1.
DR Pfam; PF00550; pp-binding; 4.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 4.
DR PROSITE; PS00455; AMP_BINDING; 4.
DR PROSITE; PS00075; ACP_DOMAIN; 4.
KW Ligase; Antibiotic biosynthesis; Multifunctional enzyme;
KW Repeat; Phosphopantetheine; Hydrolase.
FT INIT_MET 0
FT REPEAT 466 1043 DOMAIN 1 (PROLINE-ACTIVATING).
FT REPEAT 1521 2080 DOMAIN 2 (VALINE-ACTIVATING).
FT REPEAT 2538 3135 DOMAIN 3 (ORNITHINE-ACTIVATING).
FT REPEAT 3591 4173 DOMAIN 4 (LEUCINE-ACTIVATING).
FT DOMAIN 975 1042 ACYL CARRIER (ACP).
FT DOMAIN 2011 2078 ACYL CARRIER (ACP).
FT DOMAIN 3057 3124 ACYL CARRIER (ACP).
FT DOMAIN 4095 4162 ACYL CARRIER (ACP).
FT BINDING 1005 1005 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 2041 2041 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 3087 3087 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 4125 4125 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT CONFLICT 274 274 H -> D (IN REF. 2).
FT CONFLICT 418 418 K -> R (IN REF. 2).
FT CONFLICT 654 664 NAVLTCTVTKKF -> TCSFDVCYQEI (IN REF. 2).
FT CONFLICT 941 946 HVRLHL -> QLPLTP (IN REF. 2).
SQ SEQUENCE 4451 AA; 510036 MW; E0029C9B51F5A4B7 CRC64;

```

Query Match 29.1%; Score 48; DB 1; Length 4451;  
Best Local Similarity 50.0%; Pred. No. 1.5e+02;  
Matches 11; Conservative 3; Mismatches 4; Indels 4; Gaps 1;

QY 8 KFIGIT----ELGGHEIKKVLV 25  
: ||||| :  
Db 4110 ELIGITDNEFFELGGHSLKATLL 4131

Search completed: April 14, 2001, 10:17:38  
Job time: 340 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2001, 10:09:35 ; Search time 77.31 Seconds  
(without alignments)  
46.998 Million cell updates/sec

Title: US-09-362-731-1  
Perfect score: 165  
Sequence: 1 QYIKANSKFIGITELGGHEIKKVLVPGCHGS 31

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL\_15:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_unclassified:\*
- 13: sp\_vertebrate:\*
- 14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	74	44.8	145	5	O96430
2	73	44.2	135	5	O9T222
3	54	32.7	1333	5	Q24262
4	50.5	30.6	60	7	Q31585
5	50.5	30.6	71	7	Q9XRJ9
6	50.5	30.6	244	7	Q31590
7	50.5	30.6	433	2	Q9PAB9
8	50	30.3	697	2	Q45419
9	49.5	30.0	67	7	Q31578
10	49	29.7	354	4	Q9UPT4
11	49	29.7	604	2	Q9RTX4
12	48.5	29.4	237	2	Q9RW82
13	48	29.1	440	11	Q35849
14	48	29.1	4450	2	Q44928
15	47.5	28.8	149	7	Q31495
16	47.5	28.8	379	5	Q23918
17	47.5	28.8	2467	14	Q9J1K2
18	47	28.5	147	10	O92VX0
19	47	28.5	162	10	O9LXF9

20	47	28.5	278	5	Q9VD00
21	47	28.5	395	5	Q9VK80
22	47	28.5	438	10	Q9MAH3
23	47	28.5	440	6	Q9WZ04
24	47	28.5	458	5	O44506
25	47	28.5	1101	10	Q9ZQ77
26	47	28.5	2138	5	Q9XZE3
27	46.5	28.2	67	7	Q31577
28	46.5	28.2	67	7	Q31581
29	46.5	28.2	67	7	Q31582
30	46.5	28.2	245	7	Q31591
31	46.5	28.2	937	10	Q9LRN1
32	46.5	28.2	1038	5	Q9V8P9
33	46.5	28.2	1087	5	Q25961
34	46.5	28.2	1694	5	Q9TZT5
35	46.5	28.2	1694	5	Q9NXX1
36	46.5	28.2	1704	5	Q9TZT4
37	46.5	28.2	1720	5	Q25922
38	46	27.9	124	2	Q9ZB68
39	46	27.9	210	10	Q39969
40	46	27.9	226	4	Q9UJG9
41	46	27.9	257	13	Q9YGJ9
42	46	27.9	466	2	Q51693
43	46	27.9	466	2	Q9RMX7
44	46	27.9	471	5	Q9Y1X6
45	46	27.9	487	2	Q9R9E4

#### ALIGNMENTS

RESULT 1  
O96430 PRELIMINARY; PRT; 145 AA.  
AC O96430;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE GROUP 2 ALLERGEN EUR M 2 0101.  
GN EUR M 2 0101

OS Euroglyphus maynei (House-dust mite).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
OC Acariformes; Sarcoptiformes; Astigmata; Analgoidea; Pyroglyphidae;  
OC Euroglyphus.  
OX NCBI\_TaxID=6958;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Smith W., Hart B.J., Thomas W.R.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF047613; AAC82349.1; -  
DR HSSP: P49278; 1A9V  
DR INTERPRO: IPR003172; -  
DR PFAM: PF02221; EL\_DerP2\_DerF2; 1.  
SQ SEQUENCE 145 AA; 15747 MW; 6655B16C8503A565 CRC64;

Query Match 44.8%; Score 74; DB 5; Length 145;  
Best Local Similarity 59.1%; Pred. No. 0.0044;  
Matches 13; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 10 IGITELGGHEIKKVLVPGCHGS 31  
Db 19 VDIKDCANHEIKKVMVPGCKGS 40

RESULT 2  
O9T222 PRELIMINARY; PRT; 135 AA.  
ID O9T222  
AC O9T222;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE GROUP 2 ALLERGEN EUR M 2 0102 (FRAGMENT).

```

GN EUR M 2 0102.
OS Euroglyphus maynei (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcotiformes; Astigmata; Analgoidea; Pyroglyphidae;
OC Euroglyphus.
OX NCBI_TaxID=6958;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith W., Hart B.J., Thomas W.R.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF047614; AAC62350.1; -.
DR HSSP: P49278; 1A9V.
DR INTERPRO: IPR003172; -.
DR PFAM: PF02221; EL_Derp2_DerF2; 1.
FT NON_TER 1
SQ SEQUENCE 135 AA; 14555 MW; E15E1FA480B8DF16 CRC64;

Query Match 44.2%; Score 73; DB 5; Length 135;
Best Local Similarity 54.5%; Pred. No. 0.0057;
Matches 12; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 10 IGITELGGHEIKKVLVPGCHGS 31
Db : : : : ||||| : |||||
9 VDVKDCANHEIKKVMVPGCKGS 30

RESULT 3
Q24262 PRELIMINARY; PRT; 1333 AA.
ID Q24262;
AC Q24262;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE BLASTOPIA POLYPROTEIN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=OREGON R;
RC MEDLINE=94333069; PubMed=8055714;
RA Frommer G., Schuh R., Jackle H.;
RT "Localized expression of a novel microRNA-like element in the
RT blastoderm of Drosophila melanogaster is dependent on the anterior
RT morphogen bicoid.";
RL Chromosome 103:82-89(1994).
DR EMBL: Z27119; CAA81643.1; -.
DR FLYBASE: FBgn0002745; microRNA.
DR FLYBASE: FBgn0014947; blastopia.
DR INTERPRO: IPR000477; -.
DR INTERPRO: IPR001584; -.
DR INTERPRO: IPR001878; -.
DR INTERPRO: IPR001969; -.
DR INTERPRO: IPR001995; -.
DR PFAM: PF00078; rvt; 1.
DR PFAM: PF00098; zf-CCHC; 2.
DR PFAM: PF00665; rve; 1.
DR PRINTS: PR00939; C2HCZNFINGER.
DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_1.
KW Polyprotein.
SQ SEQUENCE 1333 AA; 152451 MW; 2383E01108216E36 CRC64;

Query Match 32.7%; Score 54; DB 5; Length 1333;
Best Local Similarity 35.9%; Pred. No. 43;
Matches 14; Conservative 7; Mismatches 10; Indels 8; Gaps 1;

QY 1 QYIKANSFIGITEL-----GGHEIKKVLVPGCHGS 31
Db : : : : ||||| : |||||
127 KYVQARSRMIGSAELFSECVSGVTELKELLIEFSGS 165

```

```

RESULT 4
Q31585 PRELIMINARY; PRT; 60 AA.
ID Q31585;
AC Q31585;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE (DB02) MHC CLASS II BETA 1 (FRAGMENT).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RA Grimholt U., Olsaker I., Vries Lindstrom C., Lie O.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL: L24953; AAA49597.1; -.
DR INTERPRO: IPR000353; -.
DR PFAM: PF00969; MHC_II_beta; 1.
KW MHC.
FT NON_TER 1
FT NON_TER 60
SQ SEQUENCE 60 AA; 6776 MW; 5D4F3449060940E2 CRC64; ✓

Query Match 30.6%; Score 50.5; DB 7; Length 60;
Best Local Similarity 57.9%; Pred. No. 4;
Matches 11; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITELG 16
Db : : : : |||||
16 EYIRFNSTVGKFGVGYTELG 34

RESULT 5
Q9XRJ9 PRELIMINARY; PRT; 71 AA.
ID Q9XRJ9;
AC Q9XRJ9;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE MHC CLASS II BETA 1 (FRAGMENT).
GN SANA.
OS Salvelinus namaycush (lake trout).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salvelinus.
OX NCBI_TaxID=8040;
RN [1]
RP SEQUENCE FROM N.A.
RA Dorschner M.O., Duris T., Phillips R.B.;
RL "Diversity of a Lake Trout Mhc class II Gene.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF130026; AAD20889.1; -.
DR INTERPRO: IPR000353; -.
DR PFAM: PF00969; MHC_II_beta; 1.
FT NON_TER 1
FT NON_TER 71
SQ SEQUENCE 71 AA; 8002 MW; E3095286582A9F2D CRC64;

Query Match 30.6%; Score 50.5; DB 7; Length 71;
Best Local Similarity 57.9%; Pred. No. 4.9;
Matches 11; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITELG 16
Db : : : : |||||
14 EYIRFNSTVGKFGVGYTELG 32

```

## RESULT 6

Q31590 ID Q31590 PRELIMINARY; PRT; 244 AA.  
 AC Q31590;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE MHC CLASS II.  
 OS Salmo salar (Atlantic salmon).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.  
 OX NCBI\_TaxID=8030;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LEUKOCYTES;  
 RA Hordvik I., Grimholt U., Fosse V.M., Lie Y., Endresen C.;  
 RL Immunogenetics 0:0-0(0).  
 DR EMBL; X70166; CAA49725.1; -;  
 DR INTERPRO; IPR000353; -;  
 DR INTERPRO; IPR003006; -;  
 DR PFAM; PF00047; 1g; 1.  
 DR PFAM; PF00969; MHC\_II\_beta; 1.  
 DR PRODOM; PD000328; -; 1.  
 KW MHC.  
 SQ SEQUENCE 244 AA; 27449 MW; 496CB9EA9D73765C CRC64;

Query Match 30.6%; Score 50.5; DB 7; Length 244;  
 Best Local Similarity 57.9%; Pred. No. 20;  
 Matches 11; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 1 QYIKANS---KFIGITELG 16  
 :||: || ||:|||||  
 Db 51 EYIRFNSTVGKFGVYTELG 69

## RESULT 7

Q9PAB9 ID Q9PAB9 PRELIMINARY; PRT; 433 AA.  
 AC Q9PAB9;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE PHOSPHODIESTERASE-NUCLEOTIDE PYROPHOSPHATASE PRECURSOR.  
 GN XF2599.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Xylella.  
 OX NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9ASC;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,

RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";  
 RL Nature 406:151-157(2000).  
 DR EMBL; AE004067; AAF85396.1; -;  
 DR INTERPRO; IPR002591; -;  
 DR PFAM; PF01663; phosphodiester; 1.  
 SQ SEQUENCE 433 AA; 47292 MW; 774F3D3E31BA52C4 CRC64;

Query Match 30.6%; Score 50.5; DB 2; Length 433;  
 Best Local Similarity 32.3%; Pred. No. 38;  
 Matches 10; Conservative 10; Mismatches 10; Indels 1; Gaps 1;

QY 2 YIKANSKFIGITELGGHEIK-KVLVPGCHGS 31  
 :::: : |||| | ||| | : : | :  
 Db 273 HWSDGQVLGITPLPGHESKVEAMLLGAHAT 303

## RESULT 8

Q45419 ID Q45419 PRELIMINARY; PRT; 697 AA.  
 AC Q45419;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE PUTATIVE TRANSCRIPTIONAL REGULATOR.  
 GN MTLR.  
 OS Bacillus stearothermophilus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1422;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 7954;  
 RX MEDLINE=96421984; PubMed=8824601;  
 RA Henstra S.A., Tolner B., ten Hoeve Duurkens R.H., Konings W.N.,  
 RA Robillard G.T.;  
 RT "Cloning, expression, and isolation of the mannitol transport protein  
 from the thermophilic bacterium Bacillus stearothermophilus.";  
 RL J. Bacteriol. 178:5586-5591(1996).  
 DR EMBL; U18943; AAC44464.1; -;  
 DR INTERPRO; IPR001550; -;  
 DR PFAM; PF00874; BglG\_antitermin; 2.  
 DR PRODOM; PD004151; -; 1.  
 SQ SEQUENCE 697 AA; 79242 MW; BB052EC2A5C7F94F CRC64;

Query Match 30.3%; Score 50; DB 2; Length 697;  
 Best Local Similarity 44.4%; Pred. No. 78;  
 Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 12 ITELGGHEIKKVLVPGCH 29  
 : : || ||: ||  
 Db 543 LEQVQGHSTKEVLAEACH 560

## RESULT 9

Q31578 ID Q31578 PRELIMINARY; PRT; 67 AA.  
 AC Q31578;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 DE (DB03) MHC CLASS II BETA 1 (FRAGMENT).  
 OS Salmo salar (Atlantic salmon).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;



```

RESULT 13
O35849 ID O35849 PRELIMINARY; PRT; 440 AA.
AC AT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE LECITHIN:CHOLESTEROL ACYLTRANSFERASE (EC 2.3.1.43)
DE (PHOSPHATIDYLCHOLINE--STEROL O-ACYLTRANSFERASE)
DE (LECITHIN--CHOLESTEROL ACYLTRANSFERASE) (LCAT)
DE (PHOSPHOLIPID--CHOLESTEROL ACYLTRANSFERASE).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY, AND WISTAR; TISSUE=LIVER;
RX MEDLINE=97363611; PubMed=9219904;
RA Wang J., Gebre A.K., Anderson R.A., Parks J.S.;
RT "Cloning and in vitro expression of rat lecithin:cholesterol
RT acyltransferase.";
RL Biochim. Biophys. Acta 1346:207-211(1997).
CC -I- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + STEROL = STEROL ESTER +
CC 1-ACYLGLYCEROPHOSPHOCHOLINE.
DR EMBL; U62803; AAB65771.1; -
DR INTERPRO; IPR000379; -
DR INTERPRO; IPR000734; -
DR PROSITE; PS00120; LIPASE_SER; 1.
KW Transferase; Acyltransferase.
SQ SEQUENCE 440 AA; 49882 MW; 836BB2D589f72B63 CRC64;

Query Match 29.1%; Score 48; DB 11; Length 440;
Best Local Similarity 50.0%; Pred. No. 90;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 14 ELGGHEIKKVLVPGCHGS 31
II I :|||| I:
DB 41 ELSNHRPVLVPGCMGN 58

RESULT 14
O44928 ID O44928 PRELIMINARY; PRT; 4450 AA.
AC AT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE GRAMICIDIN S SYNTHETASE 2.
DE GRS2.
GN Bacillus brevis.
OS Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Brevibacillus.
OX NCBI_TaxID=1393;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NAGANO;
RX MEDLINE=95122465; PubMed=7822255;
RA Saito F., Hori K., Kanda M., Kurotsu T., Saito Y.;
RT "Entire nucleotide sequence for Bacillus brevis Nagano GRS2 gene
RT encoding gramicidin S synthetase 2: a multifunctional peptide
RT synthetase.";
RL J. Biochem. 116:357-367(1994).
DR EMBL; D29676; BAA06146.1; -
DR HSSP; P14687; 1AMU.
DR INTERPRO; IPR000255; -
DR INTERPRO; IPR000379; -
DR INTERPRO; IPR000873; -
DR INTERPRO; IPR001031; -
DR INTERPRO; IPR001242; -
DR INTERPRO; IPR002086; -
PFAM; PF00501; AMP-binding; 4.

```

```

DR PFAM; PF00550; PP-binding; 4.
DR PFAM; PF00668; DUF4; 4.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_4.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_2.
DR PROSITE; PS00455; AMP_BINDING; 4.
DR PROSITE; PS00075; ACP_DOMAIN; 4.
SQ SEQUENCE 4450 AA; 508681 MW; F3197E77BF69316D CRC64;

Query Match 29.1%; Score 48; DB 2; Length 4450;
Best Local Similarity 50.0%; Pred. No. 13e+03;
Matches 11; Conservative 3; Mismatches 4; Indels 4; Gaps 1;

OY 8 KFIGIT---ELGGHEIKKVLV 25
: |||| | |||| I:
DB 4109 ELGITDNFFELGGHSLKATLL 4130

RESULT 15
O31495 ID O31495 PRELIMINARY; PRT; 149 AA.
AC AT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE MHC CLASS II BETA (FRAGMENT)
OS Oncorhynchus masou (Cherry salmon) (Masu salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8020;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BLOOD;
RX MEDLINE=96186525; PubMed=8606054;
RA Miller K.M., Withler R.E.;
RT "Sequence analysis of a polymorphic Mhc class II gene in Pacific
RT salmon.";
RL Immunogenetics 43:337-351(1996).
DR EMBL; U34700; AAB01698.1; -
DR INTERPRO; IPR000353; -
DR INTERPRO; IPR003006; -
DR PFAM; PF00047; ig; 1.
DR PFAM; PF00969; MHC_II_beta; 1.
KW MHC.
FT NON_TER 1
FT NON_TER 149
SQ SEQUENCE 149 AA; 17003 MW; 5B56B3C2A74666EE CRC64;

Query Match 28.8%; Score 47.5; DB 7; Length 149;
Best Local Similarity 52.6%; Pred. No. 31;
Matches 10; Conservative 4; Mismatches 2; Indels 3; Gaps 1;

OY 1 QYIKANS---KFIGITELG 16
: ||: | :|: |||
DB 14 BYIRFNSTVGRFVGYTELG 32

```

Search completed: April 14, 2001, 10:13:06  
Job time: 211 sec